Sebastian Volz
Editor

TOPICS IN APPLIED PHYSICS 118

Thermal Nanosystems and Nanomaterials

Springer

Topics in Applied Physics
Volume 118

Available online at
SpringerLink.com

Topics in Applied Physics is part of the SpringerLink service. For all customers with standing orders for Topics in Applied Physics we offer the full text in electronic form via SpringerLink free of charge. Please contact your librarian who can receive a password for free access to the full articles by registration at:

springerlink.com → Orders

If you do not have a standing order you can nevertheless browse through the table of contents of the volumes and the abstracts of each article at:

springerlink.com → Browse Publications

Topics in Applied Physics

Topics in Applied Physics is a well-established series of review books, each of which presents a comprehensive survey of a selected topic within the broad area of applied physics. Edited and written by leading research scientists in the field concerned, each volume contains review contributions covering the various aspects of the topic. Together these provide an overview of the state of the art in the respective field, extending from an introduction to the subject right up to the frontiers of contemporary research.
Topics in Applied Physics is addressed to all scientists at universities and in industry who wish to obtain an overview and to keep abreast of advances in applied physics. The series also provides easy but comprehensive access to the fields for newcomers starting research.
Contributions are specially commissioned. The Managing Editors are open to any suggestions for topics coming from the community of applied physicists no matter what the field and encourage prospective editors to approach them with ideas.

Managing Editor

Dr. Claus E. Ascheron

Springer-Verlag GmbH
Tiergartenstr. 17
69121 Heidelberg
Germany
Email: claus.ascheron@springer.com

Assistant Editor

Adelheid H. Duhm

Springer-Verlag GmbH
Tiergartenstr. 17
69121 Heidelberg
Germany
Email: adelheid.duhm@springer.com

For further volumes:
http://www.springer.com/series/560

Sebastian Volz

Editor

Thermal Nanosystems and Nanomaterials

With 261 Figures

Springer

Editor
Dr. Sebastian Volz
Laboratoir d'Energétique Moléculaire et Macroscopique Combustion
Ecole Central Paris UPR CNRS 288
Grande Voie des Vignes
92295 Châtenay Malabry, France
E-mail: volz@em2c.ecp.fr

ISSN 0303-4216
ISBN 978-3-642-04257-7 e-ISBN 978-3-642-04258-4
DOI 10.1007/978-3-642-04258-4
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009942090

© Springer-Verlag Berlin Heidelberg 2009
This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Five years ago, a first book entitled *Microscale and Nanoscale Heat Transfer* was put together, and eventually published in 2007 as number 107 in the Springer TAP series. The aim was to bring together a group of scientists with a common interest in heat transfer problems on the micro- and nanoscales. Since then, it has become clear that these problems constitute a key feature of the nanoscience adventure. Apart from the fact that energy has become a major theme today, for which solutions are required across the board, including in the area of nanotechnology, nanoscale heat transfer is now a major issue for very high level international research groups who are moving beyond the limits set by their predecessors and whose reports are now being published in high-profile, wide-readership journals such as Nature and Science. Knowledge in this area has thus progressed, and its applications are ever more diverse.

Although the present manuscript has been put together by a similar scientific community to the one that produced the first volume, i.e., a research group under the aegis of the French National Research Administration (CNRS), referring essentially to the departments of engineering and information science, the book is not merely an extension of the first volume, since a large part of it is devoted to a whole range of new applications.

The field of applications is divided into two main parts. The first corresponds to Part I of the book and concerns nanomaterials and their heat transfer properties. This part is itself divided into two themes. The first, Chaps. 1–3, is a somewhat theoretical review of the physics of nanostructures, while the second, Chaps. 4–8, deals with the effective properties of composites. Part II concerns microsystems and three types of application: thermoelectric energy conversion systems in Chap. 9, in vitro and in vivo biological systems in Chaps. 10 and 11, respectively, and microelectronic systems in Chap. 12.

Following this is a third and final part relating to advanced thermal measurement techniques. Some of these are recent developments of methods already introduced in the previous volume, while others correspond to fundamentally new systems. This part contains two chapters devoted to optical metrology (Chaps. 13 and 14), three describing different forms of local probe microscopy (Chaps. 15–17), and finally a chapter on low-temperature thermometry (Chap. 18).

This book brings together a quite remarkable overview of the state of the art, producing a continuous spectrum of subjects that broadly covers all the various fields of application of nanoscale heat transfer from solid state physics to biology. Such a perspective should be useful for doctoral students wishing to obtain their own general awareness of the specific themes discussed here. But it should also provide a way for practising research scientists to enter this particularly rich area of investigation, deepening and broadening their own skills or making their own contribution to the field.

Acknowledgments

The authors would like to acknowledge the financial and administrative support of the Institut des Sciences et Technologies de l'Information et de l'Ingénierie (INST2I) of the Centre National de la Recherche Scientifique (CNRS) and the Societe Française de Thermique. My sincere thanks go to all the authors who have contributed to this book for the very high standard of their work. Most are long-standing friends. It is my hope that their scientific knowledge and understanding will thereby be disseminated to the benefit of the international scientific community.

Tokyo, March 2009 *Sebastian Volz*

Contents

Part I Nanomaterials

1 Introduction 3
Sebastian Volz
1.1 Nanostructures 3
1.2 Scientific and Technological Stakes 5
1.3 Physical Mechanisms 8
1.3.1 Rarefaction. Surface Reflection and Transmission at Interfaces 9
1.3.2 Confinement 10
1.3.3 Densities of States and Dimensionality 12
1.3.4 Non-Fourier Effects and Thermal Conductivity 13
1.4 Conclusion 15
References 15

2 Nanostructures 17
Patrice Chantrenne, Karl Joulain, and David Lacroix
2.1 Introduction 17
2.2 Modelling Heat Transfer 18
2.2.1 The Physics of Phonon Transfer 18
2.2.2 Semi-Analytic Models 27
2.3 Nanofilms, Nanowires, and Nanotubes 31
2.3.1 Deterministic Model: BTE and the Discrete Ordinate Method 31
2.3.2 Statistical Model: BTE and the Monte Carlo Method 33
2.3.3 Mechanical Model: Molecular Dynamics 40
2.4 Comparison and Limitations of the Models 49
2.4.1 Examples of Confinement in a Nanofilm 50
2.4.2 Examples of Confinement in a Nanowire and a Nanotube 54

2.5 Conclusion . . . 57
Appendix: Measuring Thermal Properties . . . 58
References . . . 59

3 Green's Function Methods for Phonon Transport Through Nano-Contacts . . . 63
Natalio Mingo
3.1 Introduction to Green's Functions for Lattice Thermal Transport . . . 63
3.2 The Harmonic Problem . . . 65
3.2.1 Dynamics of Non-Periodic Systems . . . 65
3.2.2 The Heat Current . . . 68
3.2.3 Different Formulas for the Transmission . . . 71
3.2.4 Weak Coupling Limit: The Low Temperature Thermal Conductance of a Weak Junction . . . 75
3.2.5 Upper Limits to Thermal Conductance, Entropy Flow, and Information Rates . . . 77
3.3 The Anharmonic Problem . . . 82
3.3.1 Many-Body Hamiltonian . . . 82
3.3.2 The Heat Current . . . 83
3.3.3 Computing the Interacting Phonon Green Functions . . . 84
3.3.4 Another Formula for the Heat Current . . . 89
3.3.5 Can We 'See' the Phonon Current? . . . 90
3.4 Concluding Remarks . . . 91
References . . . 93

4 Macroscopic Conduction Models by Volume Averaging for Two-Phase Systems . . . 95
Benoît Goyeau
4.1 Introduction . . . 95
4.2 Local Volume Averages . . . 96
4.3 Averaged Equations . . . 97
4.3.1 Local Thermal Equilibrium and the Single-Equation Model . . . 99
4.3.2 Deviation Equations . . . 100
4.3.3 Closure Problem . . . 103
4.3.4 Closed Form . . . 104
4.3.5 Local Thermal Non-Equilibrium . . . 105
References . . . 105

5 Heat Conduction in Composites . . . 107
Jean-Yves Duquesne
5.1 Microcomposites and Effective Media . . . 107
5.1.1 Taking Averages . . . 107
5.1.2 Particle Shell . . . 110
5.1.3 Experimental Examples . . . 114
5.2 Nanocomposites and Phonon Scattering . . . 114
5.2.1 Limitations of Effective Medium Theories . . . 114
5.2.2 Kinetic Theory of Heat Transport in Solids . . . 115

5.2.3 Phonon Scattering by Particles 118
5.2.4 Example of a Pure Bulk Material 118
5.2.5 Example of a Disordered Alloy 120
5.3 Conclusion 122
Appendix A. Demonstration of (5.7) 122
Appendix B. Effective Medium and Interface Resistance 123
Appendix C. Calculation Parameters for Scattering by Particles 125
References 126

6 Optical Generation and Detection of Heat Exchanges in Metal–Dielectric Nanocomposites 127
Bruno Palpant
6.1 Optical Properties of Noble Metal Nanoparticles and Nanocomposite Media 128
6.1.1 Dielectric Function of Noble Metals 128
6.1.2 Optical Response of Nanocomposite Media 131
6.2 Thermo-Optical Response 132
6.2.1 Noble Metals 132
6.2.2 Nanocomposites 133
6.2.3 Calculation 134
6.3 Heat Exchange Dynamics 136
6.3.1 Athermal Regime and the Boltzmann Equation 136
6.3.2 Thermal Regime and the Three-Temperature Model 139
References 147

7 Mie Theory and the Discrete Dipole Approximation. Calculating Radiative Properties of Particulate Media, with Application to Nanostructured Materials 151
Franck Enguehard
7.1 Introduction 151
7.2 Absorption and Scattering by a Particle of Arbitrary Shape and by a Population of Such Particles 153
7.2.1 Incident Electromagnetic Field, Poynting Vector, and Associated Power 153
7.2.2 Electromagnetic Fields Within and Scattered by the Particle 154
7.2.3 Extinction, Absorbed, and Scattered Power 155
7.2.4 Expressing the Extinction and Scattered Powers in Terms of the Incident and Scattered Electric Fields . . . 157
7.2.5 Extinction, Absorption, and Scattering Cross-Sections. Associated Efficiencies and Scattering Phase Function . . . 159
7.2.6 Directions of Propagation and Polarisation 160
7.2.7 Radiative Properties of a Population of Particles 162
7.3 Mie Theory 163
7.3.1 Analytic Solution to Mie's Electromagnetic Problem 163
7.3.2 Extinction and Scattering Cross-Sections. Scattering Phase Function 165

7.3.3 A Special Case: Rayleigh Scattering . . . 167
7.3.4 Numerical Considerations . . . 169
7.3.5 Radiative Response of a Population of Spherical Particles . . . 170
7.3.6 Application of Mie Theory to the Radiative Response of a Cloud . . . 173
7.4 Discrete Dipole Approximation (DDA) . . . 175
7.4.1 The Theory of the DDA . . . 176
7.4.2 Models for Polarisability . . . 187
7.4.3 Applying the Discrete Dipole Approximation . . . 191
7.5 Summary . . . 205
Appendix: Analytical Solution of Mie's Electromagnetic Problem . . . 205
References . . . 211

8 Thermal Conductivity of Nanofluids . . . 213
Pawel Keblinski
8.1 Introduction . . . 213
8.2 Excitement, Controversy, and New Physics . . . 214
8.2.1 Brownian Motion . . . 214
8.2.2 Interfacial Liquid Layer . . . 215
8.2.3 Interfacial Thermal Resistance . . . 216
8.2.4 Near Field Radiation . . . 216
8.2.5 Particle Clustering . . . 217
8.3 Discussion . . . 218
References . . . 219

Part II Nanosystems

9 Nanoengineered Materials for Thermoelectric Energy Conversion . . 225
Ali Shakouri and Mona Zebarjadi
9.1 Introduction . . . 225
9.2 Thermoelectric Energy Conversion . . . 226
9.3 Theoretical Modelling . . . 230
9.3.1 Boltzmann Transport and Thermoelectric Effects . . . 230
9.3.2 Theory of Thermoelectric Transport in Multilayers and Superlattices . . . 233
9.3.3 Monte Carlo Simulation of Electron Transport in Thermoelectric Layers . . . 235
9.3.4 Non-Equilibrium Green Function for Thermoelectric Transport . . . 236
9.3.5 Phonon Transport . . . 237
9.3.6 Thermoelectric Transport in Strongly Correlated Systems . . . 238
9.3.7 Wave or Particle Picture for Electrons and Phonons? . . . 239
9.3.8 Why Is There a Trade-off Between Electrical Conductivity and Seebeck Coefficient? . . . 240
9.3.9 Low-Dimensional Thermoelectrics . . . 241

9.4 Thermionic Energy Conversion 244
9.4.1 Vacuum Thermionic Energy Conversion 244
9.4.2 Nanometer Gaps and Thermotunneling 244
9.4.3 Inverse Nottingham Effect and Carbon Nanotube Emitters 245
9.4.4 Single Barrier Solid-State Thermionic Energy Conversion 246
9.4.5 Multilayer Solid-State Thermionic Energy Conversion . . 247
9.4.6 Conservation of Transverse Momentum in Thermionic Emission 248
9.4.7 Electron Group Velocity and the Electronic Density of States 249
9.4.8 Reversible Thermoelectrics 251
9.5 Reduction of Phonon Thermal Conductivity 251
9.5.1 Thermal Conductivity of Superlattices 252
9.5.2 Thermal Conductivity of Nanowires 255
9.6 Applications 255
9.6.1 Heterostructure Integrated Thermoelectric/Thermionic Microrefrigerators on a Chip 255
9.6.2 SiGe and SiGeC Superlattice Optimization 262
9.6.3 Potential Metal/Semiconductor Heterostructure Systems . 264
9.6.4 InGaAlAs Embedded with ErAs Nanoparticles 265
9.6.5 Metal/Semiconductor Multilayers Based on Nitrides 270
9.7 Scaling up Production 271
9.7.1 Thin-Film Power Generation Modules 272
9.7.2 Optoelectronic and Electronic Applications 273
9.8 System Requirements for Power Generation 275
9.9 Graded Materials 277
9.10 Characterization Techniques 278
9.10.1 Cross-Plane Seebeck Measurement 279
9.10.2 Transient *ZT* Measurement 280
9.10.3 Suspended Heater and Nanowire Characterization 281
9.11 Thermoelectric/Thermionic vs. Thermophotovoltaics 282
9.12 Ballistic Electron and Phonon Transport Effects 283
9.13 Nonlinear Thermoelectric Effects 285
9.14 A Refrigerator Without the Hot Side 286
9.15 Conclusion 287
References 287

10 Molecular Probes for Thermometry in Microfluidic Devices 301
Charlie Gosse, Christian Bergaud, and Peter Löw
10.1 Microlaboratories and Heat Transfer Issues 301
10.1.1 Historical Context 301
10.1.2 Electrokinetic Separation 302
10.1.3 DNA Amplification by PCR 303
10.1.4 Thermodynamic and Kinetic Measurements 305

10.2 Microfluidics and Thermometry 308
10.2.1 Electrical Methods 308
10.2.2 Non-Spectroscopic Optical Methods 309
10.2.3 Molecular-Probe-Related Methods 310
10.3 Thermosensitive Materials 311
10.3.1 Liquid Crystals 311
10.3.2 Polymers 315
10.3.3 Phospholipid Membranes 316
10.4 Kinetic Fluorescent Probes 317
10.4.1 Intramolecular Charge Transfer in Organic Molecules . . . 319
10.4.2 Charge Transfer in Organometallic Complexes 320
10.4.3 Excimer Formation 322
10.4.4 Delayed Fluorescence 323
10.5 Thermodynamic Fluorescent Probes 324
10.5.1 Isomerisation Between Species in Their Ground State. General Features 325
10.5.2 Folding of Nucleic Acid Structures 326
10.5.3 Chromophore Complexation by Cyclodextrins 327
10.5.4 Acid–Base Reactions 328
10.5.5 Modification of the Coordination Sphere of Metallic Ions 329
10.5.6 Thermalisation Between Excited States. General Features and Examples 330
10.6 Procedures for Fluorescence Microscopy 331
10.6.1 Single-Wavelength Intensity Measurement 331
10.6.2 Ratiometric Intensity Measurement 332
10.6.3 Lifetime Measurement 333
10.7 Other Forms of Spectroscopy for Probing Thermodynamic Equilibria 333
10.7.1 Raman Spectroscopy 334
10.7.2 Nuclear Magnetic Resonance 334
10.8 Conclusion and Prospects 335
10.8.1 Fluorescent Probes 336
10.8.2 Microscopic Techniques 336
References 337

11 Cell Targeting and Magnetically Induced Hyperthermia 343
Etienne Duguet, Lucile Hardel, and Sébastien Vasseur
11.1 Introduction 343
11.1.1 Nanomedicine. An Application of Nanoscience and Nanotechnology 343
11.1.2 An Incomplete and Complex Set of Requirements 344
11.2 In Vivo Applications of Nanoparticles 344
11.2.1 Nanoparticles in the Blood Compartment 344
11.2.2 Designing Particles with Extended Vascular Lifetime 345
11.2.3 Active Targeting by Coupling with Molecular Recognition Ligands 346

11.2.4 Alternatives to Active Targeting 347
11.2.5 Overview of Commercially Available and Forthcoming Formulations 348
11.2.6 Relative Importance of Toxicity 352
11.3 Magnetically Induced Hyperthermia 353
11.3.1 Therapeutic Advantages of Heat 353
11.3.2 Different Methods and Their Limitations 353
11.3.3 Mechanisms for Induction Losses in Magnetic Materials . 354
11.3.4 Comparing Theory and Experiment 356
11.3.5 Physiological Constraints 357
11.3.6 Some Formulations Under Development or Undergoing Clinical Assessment 359
11.4 Short and Mid-Term Prospects 360
11.4.1 Mediators 361
11.4.2 Physics of Magnetic Dissipation Phenomena and Modelling the in Vivo Temperature Distribution 362
11.4.3 Targeting Strategies 362
11.4.4 System Applying the Alternating Magnetic Field 362
References 363

12 Accounting for Heat Transfer Problems in the Semiconductor Industry 367
Christian Brylinski
12.1 Introduction 367
12.2 General Trends 368
12.2.1 Miniaturisation 368
12.2.2 Rising Frequencies 369
12.2.3 Heterogeneous Integration 370
12.3 Heat Transport 370
12.3.1 Heat Conduction 370
12.3.2 Microscopic Order in the Semiconductor 371
12.3.3 The Substrate 372
12.4 Problems and Predictions for the Main Chip Types 378
12.4.1 Components for Controlling Electrical Energy 378
12.4.2 Processor and Memory 381
12.4.3 Light-Emitting Components 383
12.4.4 Trends in Heat Transfer Features of Semiconductor Components in the Coming Decades . . 384
References 385

Part III Advanced Thermal Measurements at Nanoscales

13 Photothermal Techniques 389
Gilles Tessier
13.1 Introduction 389
13.1.1 Problems Specific to Structures Made by a Top–Down Approach 390
13.1.2 Thermoreflectance and CCD Cameras 391

13.2 Thermoreflectance Imaging . . . 392
13.2.1 The Underlying Phenomenon . . . 392
13.2.2 Measurement Methods . . . 393
13.3 Thermoreflectance Under Visible Illumination . . . 395
13.3.1 Spectroscopy of dR/dT . . . 395
13.3.2 Modelling . . . 396
13.3.3 Calibration . . . 398
13.4 Thermoreflectance Under Ultraviolet Illumination . . . 401
13.5 Thermoreflectance in the Near Infrared. Rear Face Imaging . . . 403
13.5.1 Near-Infrared Thermoreflectance with Laser Illumination . . . 403
13.5.2 Near-Infrared Thermoreflectance with Incoherent Illumination . . . 404
13.5.3 Improving Resolution with a Solid Immersion Lens . . . 405
References . . . 407

14 Thermal Microscopy with Photomultipliers and UV to IR Cameras . 411
Bernard Cretin and Benjamin Rémy
14.1 Basic Physics . . . 412
14.1.1 Radiometry . . . 412
14.1.2 Black Body Emission and Planck's Law . . . 415
14.1.3 Short Wavelength Measurements. Photon Flux . . . 417
14.1.4 Random Nature of the Photon Flux . . . 419
14.1.5 Multispectral Measurements . . . 420
14.2 Measurement by Photomultiplier and UV to NIR Camera . . . 421
14.2.1 Principle of Photomultipliers and Cameras . . . 422
14.2.2 Experimental Setup for the UV Thermal Microscope . . . 425
14.2.3 Experimental Setup for the Silicon CCD Camera . . . 430
14.3 Conclusion . . . 436
References . . . 437

15 Near-Field Optical Microscopy in the Infrared Range . . . 439
Yannick De Wilde, Paul-Arthur Lemoine, and Arthur Babuty
15.1 Introduction . . . 439
15.2 Resolution Limit in Conventional Microscopy . . . 441
15.3 Near-Field Microscopy . . . 444
15.3.1 Basic Idea . . . 444
15.3.2 Aperture SNOM . . . 445
15.3.3 Apertureless or Scattering SNOM . . . 448
15.4 Thermal Radiation STM . . . 456
15.4.1 Introduction . . . 456
15.4.2 TRSTM Setup and Operation . . . 457
15.4.3 First Example Application of TRSTM . . . 458
15.4.4 Prospects . . . 463
References . . . 465

16 PhotoThermal Induced Resonance. Application to Infrared Spectromicroscopy 469
Alexandre Dazzi
16.1 Infrared Spectroscopy and Microscopy 469
16.1.1 Optical Index and Absorption 469
16.1.2 Infrared Spectrometers 471
16.1.3 Confocal Microscopes 473
16.2 The PTIR Technique 475
16.3 Photothermoelastic Phenomena 477
16.4 Time Scales 477
16.5 Thermoelastic Deformation 481
16.6 AFM Contact Resonance Mode 482
16.7 Absorption Measurement by Contact Resonance 485
16.8 PTIR Lateral Resolution 488
16.8.1 Resolution of an Object Placed on a Surface 489
16.8.2 Resolution of a Buried Object 489
16.9 Experimental Illustration 492
16.9.1 Candida Albicans 492
16.9.2 Escherichia Coli and Its Bacteriophage T5 493
16.9.3 Ultralocalised Infrared Spectroscopy 494
16.9.4 Chemical Mapping at the Nanoscale 498
16.10 Conclusion 502
References 503

17 Scanning Thermal Microscopy with Fluorescent Nanoprobes 505
Lionel Aigouy, Benjamin Samson, Elika Saïdi, Peter Löw, Christian Bergaud, Jessica Labéguerie-Egéa, Carine Lasbrugnas, and Michel Mortier
17.1 Luminescence 505
17.1.1 Introduction to Luminescence 505
17.1.2 Effect of Temperature on Light Emission 507
17.2 Luminescent Materials Used in Thermometry 508
17.2.1 Organic Molecules. Intensity Variations 508
17.2.2 Materials Containing Rare Earth Ions 509
17.2.3 Materials Containing Transition Ions. Intensity Variations and Lifetimes 512
17.2.4 Semiconducting Quantum Dots. Intensity and Wavelength Variations 513
17.3 Development of a Scanning Fluorescent Probe for Temperature Measurements 515
17.3.1 Choice of Material. Reversibility 515
17.3.2 Making the Probes 516
17.3.3 Experimental Setup for Fluorescent SThM 518
17.4 Applications 519
17.4.1 Direct Current Measurements 520

17.4.2 Alternating Current Measurements 527
17.4.3 Measuring Tip–Sample Heat Transfer 531
17.5 Conclusion and Prospects 533
References 533

18 Heat Transfer in Low Temperature Micro- and Nanosystems 537
Olivier Bourgeois
18.1 Introduction 537
18.2 Thermal Physics at Low Temperatures 538
18.2.1 Equilibrium Thermodynamics 539
18.2.2 Quasi-Steady State Nonequilibrium Heat Transfer 544
18.3 Probing Thermal Properties by Electrical Measurements 551
18.3.1 Thermometry 551
18.3.2 Low Temperature Specific Heat Measurements at the Nanoscale 555
18.3.3 Thermal Conductance Measurements on Nanoscale Samples 559
18.4 Conclusions 565
References 566

Index 569

List of Contributors

Lionel Aigouy
Laboratoire Photons et Matière (LPEM), CNRS UPR 5, ESPCI 10 rue Vauquelin, 75231 Paris Cedex 5, France, e-mail: lionel.aigouy@espci.fr

Arthur Babuty
Institut Langevin, ESPCI ParisTech, CNRS UMR 7587, Laboratoire d'Optique Physique, 10 rue Vauquelin, 75231 Paris Cedex 05, France, e-mail: arthur.babuty@espci.fr

Christian Bergaud
Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS), CNRS and Université de Toulouse, UPS, INSA, INP, ISAE, 7 avenue du colonel Roche, 31077 Toulouse, France, e-mail: bergaud@laas.fr

Olivier Bourgeois
Institut Néel, CNRS-UJF, 25 avenue des Martyrs, 38042 Grenoble Cedex 9, France, e-mail: olivier.bourgeois@grenoble.cnrs.fr

Christian Brylinski
Laboratoire Multimatériaux et Interface, Université Claude Bernard Lyon I, Bâtiment Claude Berthollet, 22 rue Gaston Berger, 69622 Villeurbanne Cedex, e-mail: christian.brylinski@univ-lyon1.fr

Patrice Chantrenne
Centre de Thermique de Lyon (CETHIL), CNRS UMR 5008, INSA, Bâtiment Sadi Carnot, 20 Avenue A. Einstein, 69621 Villeurbanne Cedex, France, e-mail: patrice.chantrenne@insa-lyon.fr

Bernard Cretin
FEMTO-ST, C.N.R.S, U.M.R 6174, 32 Avenue de l'Observatoire, 25044 Besançon Cedex, France, e-mail: bcretin@femto-st.fr

Alexandre Dazzi
Laboratoire de Chimie Physique, Université de Paris-Sud, Bâtiment 201 P2, 91405 Orsay Cedex, France, e-mail: alexandre.dazzi@clio.u-psud.fr

Yannick De Wilde
Institut Langevin, ESPCI ParisTech, CNRS UMR 7587, Laboratoire d'Optique Physique, 10 rue Vauquelin, 75231 Paris Cedex 05, France, e-mail: yannick.dewilde@espci.fr

Etienne Duguet
Université de Bordeaux, Institut de Chimie de la Matière Condensée de Bordeaux, 87 avenue du Dr Albert Schweitzer, 33608 Pessac Cedex, France, e-mail: duguet@icmcb-bordeaux.cnrs.fr

Jean-Yves Duquesne
Institut des NanoSciences de Paris, UMR 7588, CNRS, Université Pierre et Marie Curie, 140 rue de Lourmel, 75015 Paris, France e-mail: Jean-Yves.Duquesne@insp.jussieu.fr

Franck Enguehard
CEA, Le Ripault, BP 16, 37260 Monts, France, e-mail: franck.enguehard@cea.fr

Charlie Gosse
Laboratoire de Photonique et de Nanostructures, LPN-CNRS, route de Nozay, 91460 Marcoussis, France, e-mail: charlie.gosse@lpn.cnrs.fr

Benoît Goyeau
Laboratoire EM2C, UPR CNRS 288, Ecole Centrale Paris, Grande Voie des Vignes, 92295 Châtenay-Malabry Cedex, France, e-mail: benoit.goyeau@em2c.ecp.fr

Lucile Hardel
Université de Bordeaux, Institut de Chimie de la Matière Condensée de Bordeaux, 87 avenue du Dr Albert Schweitzer, 33608 Pessac Cedex, France, e-mail: hardel@icmcb-bordeaux.cnrs.fr

Karl Joulain
Laboratoire d'études thermiques, Bâtiment de mécanique, 40 avenue du Recteur Pineau, 86022 Poiters Cedex, France, e-mail: karl.joulain@univ-poitiers.fr

Pawel Keblinski
Materials Science and Engineering Department, Rensselaer Polytechnic Institute, Troy, New York 12180, e-mail: keblip@rpi.edu

Jessica Labéguerie-Egéa
Laboratoire de Chimie de la Matière Condensée de Paris (LCMCP), CNRS UMR 7574, ENSCP, 11 rue Pierre et Marie Curie, 75005 Paris, France, e-mail: jessica-labeguerie@enscp.fr

David Lacroix
LEMTA, CNRS UMR 7563, Nancy Université - UHP Faculté des Sciences, BP 239, 54506 Vandœuvre Cedex, France, e-mail: David.Lacroix@lemta.uhp-nancy.fr

Carine Lasbrugnas
Laboratoire de Chimie de la Matière Condensée de Paris (LCMCP), CNRS UMR 7574, ENSCP, 11 rue Pierre et Marie Curie, 75005 Paris, France, e-mail: carine-lasbrugnas@enscp.fr

Paul-Arthur Lemoine
Institut Langevin, ESPCI ParisTech, CNRS UMR 7587, Laboratoire d'Optique Physique, 10 rue Vauquelin, 75231 Paris Cedex 05, France, e-mail: lemoine@optique.espci.fr

Peter Löw
Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS), CNRS and Université de Toulouse, UPS, INSA, INP, ISAE, 7 avenue du colonel Roche, 31077 Toulouse, France, e-mail: plow@laas.fr

Natalio Mingo
LITEN, CEA Grenoble, 17 rue des Martyrs, 38000 Grenoble, France,

Jack Baskin School of Engineering, University of California at Santa Cruz, Santa Cruz, California, e-mail: natalio.mingo@cea.fr

Michel Mortier
Laboratoire de Chimie de la Matière Condensée de Paris (LCMCP), CNRS UMR 7574, ENSCP, 11 rue Pierre et Marie Curie, 75005 Paris, France, e-mail: michel-mortier@enscp.fr

Bruno Palpant
Laboratoire de Photonique Quantique et Moléculaire, UMR 8537, CNRS, ENS Cachan, Ecole Centrale Paris, Grande Voie des Vignes F-92295 CHÂTENAY-MALABRY Cedex, France, e-mail: bruno.palpant@ecp.fr

Benjamin Rémy
2 LEMTA ENSEM, UMR 7563, 2 avenue de la forêt de Haye, B.P 160, 54504 Vandoeuvre-Lès-Nancy Cedex, France, e-mail: benjamin.remy@ensem.inpl-nancy.fr

Elika Saïdi
Laboratoire Photons Et Matière (LPEM), CNRS UPR 5, ESPCI 10 rue Vauquelin, 75231 Paris Cedex 5, France, e-mail: elika.saidi@espci.fr

Benjamin Samson
Laboratoire Photons Et Matière (LPEM), CNRS UPR 5, ESPCI 10 rue Vauquelin, 75231 Paris Cedex 5, France, e-mail: benjamin.samson@espci.fr

Ali Shakouri
Jack Baskin School of Engineering, University of California, Santa Cruz, CA 95064-1077, e-mail: ali@soe.ucsc.edu

Gilles Tessier
Laboratoire d'Optique Physique, 10 rue Vauquelin, 75231 Paris Cedex 05, France, e-mail: gilles.tessier@espci.fr

Sébastien Vasseur
Université de Bordeaux, Institut de Chimie de la Matière Condensée de Bordeaux, 87 avenue du Dr Albert Schweitzer, 33608 Pessac Cedex, France, e-mail: vasseur@icmcb-bordeaux.cnrs.fr

Sebastian Volz
Laboratoire d'Energétique Moléculaire et Macroscopique, Combustion, UPR288 CNRS Grande Voie des Vignes, 92295 Châtenay Malabry, France, e-mail: s.volz@ecp.fr

Mona Zebarjadi
Jack Baskin School of Engineering, University of California, Santa Cruz, CA 95064-1077, e-mail: mona@soe.ucsc.edu

Chapter 1
Introduction

Sebastian Volz

1.1 Nanostructures

Nanomaterials are defined here to be composites of entities with characteristic sizes in the range 0.1–500 nm, dilute or dense, capable of significantly modifying the properties of the matrix. These elements, called nanostructures, make up a five-letter 'alphabet': nanofilms, superlattices (Fig. 1.1a, stacks of nanofilms), nanowires (Fig. 1.1b), nanotubes (Fig. 1.1c), and nanoparticles (Fig. 1.1d).

These structures are synthesised either by relatively accessible chemical processes, e.g., electrochemistry, emulsions, milling, etc., or else by techniques involving large scale and costly equipment, e.g., molecular beam epitaxy (MBE) or focussed ion beams (FIB). Plasma deposition and chemical vapour deposition (CVD) chambers have a somewhat intermediate status, given that masking and etching may involve heavy investment when a high resolution is required.

Nanostructures have very different properties to macroscopic materials. These properties are usually related to mechanisms belonging to macroscopic physics. A nanowire can have a thermal conductivity 100 times lower than the bulk material [1], and a nanotube has higher thermal conductivity than diamond [2], if it is definable at all. This kind of extreme behaviour is an incentive to creating new composites whose properties would be modulated by varying the density, nature, and ordering of the included nanostructures.

Naturally, an isolated nano-object has very different properties from one that is included within a matrix. With the change of scale, the absence of percolation and the contact resistance between the structures and the matrix on the one hand and between the structures themselves on the other mean that intrinsic properties are not conserved. In order to control the effective properties of the resulting materials, it is thus essential to understand these resistances and the overall organisation of the constituent nano-objects.

Another approach here is simply to integrate a single nanostructure within a micro- or nanosystem to set up some function with a minimum amount of matter

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_1, © Springer-Verlag Berlin Heidelberg 2009

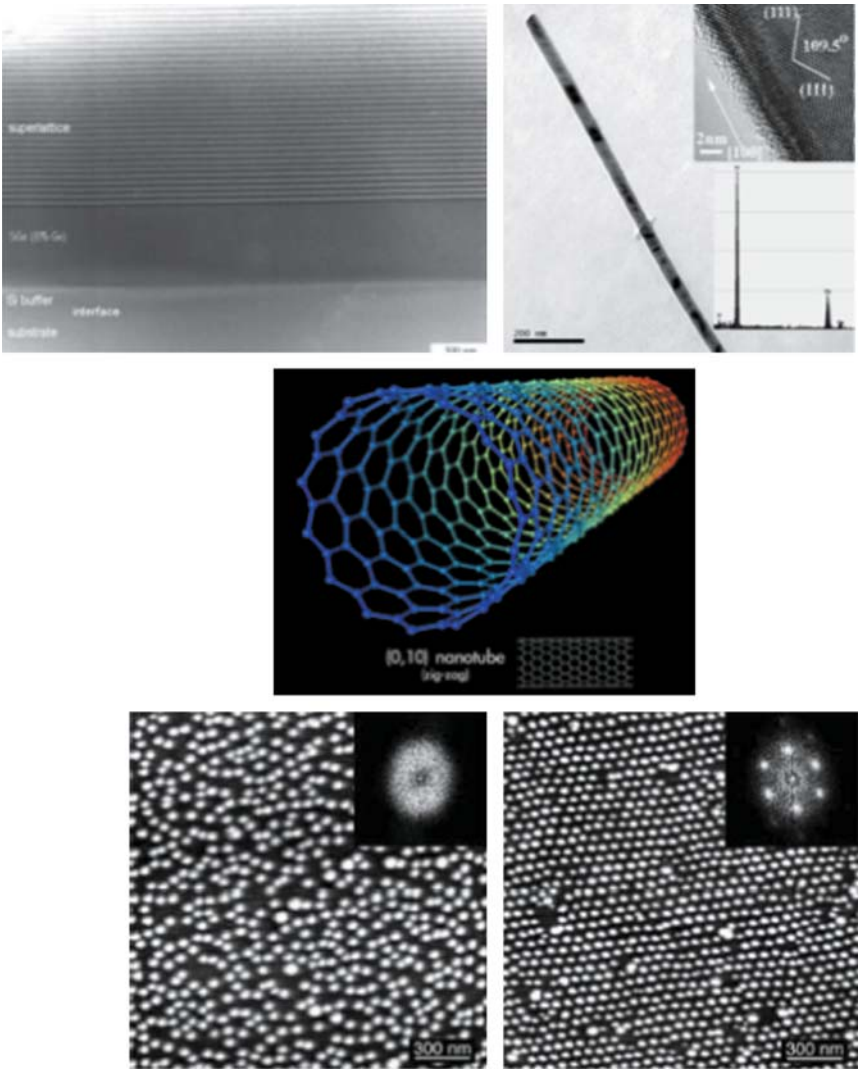

Fig. 1.1 *Top left*: Transmission electron microscope image of an Si/SiGe superlattice (Paul Schereer Institute). *Top right*: Transmission electron microscope image of a silicon nanowire. *Insert*: Magnification of the surface of the wire. *Center*: Image showing the structure of a single-wall carbon nanotube (University of Liège). *Bottom*: Atomic force microscope images of a monolayer of PbSe nanoparticles. *Inserts* show the 2D spectra of the images which reveal the ordering into hexagonal structures

and on a much reduced surface area. The original aim of nanotechnology was to assemble atoms together to build up molecules that could carry out some precise function in such a way as to save on the amount of material and ensure the sustainable development of contemporary electronics and telecommunications.

This approach was not continued for political reasons, but the pursuit of miniaturisation has nevertheless led to a drastic reduction in the dimensions of technological systems.[1] In 2008, an electronic chip with surface area around 1 cm^2 contains 250 million transistors (45 nm technology) and 6 km of interconnect tracks. In 2020, this same chip should carry as many transistors as the brain contains neurons.

In this first chapter of Part I, we present the various thermal applications of nanomaterials, then discuss the physical mechanisms underlying the novel properties of nanostructures.

1.2 Scientific and Technological Stakes

Thermal applications are currently being developed for the purposes of insulation, comfort, cooling, and also energy conversion. For example, aerogels are ultraporous media made up of fibres, themselves resulting from the coalescence of silica nanoparticles. By packing these materials in vacuum, their thermal conductivity can be made smaller than the thermal conductivity of air [3]. Figure 1.2a shows the equivalent thickness of rock wool needed to achieve the same result as an aerogel panel.

Phase-change microcapsules represent a huge market in the context of technical textiles (see Fig. 1.3). Phase-change materials provide a way of imposing a constant

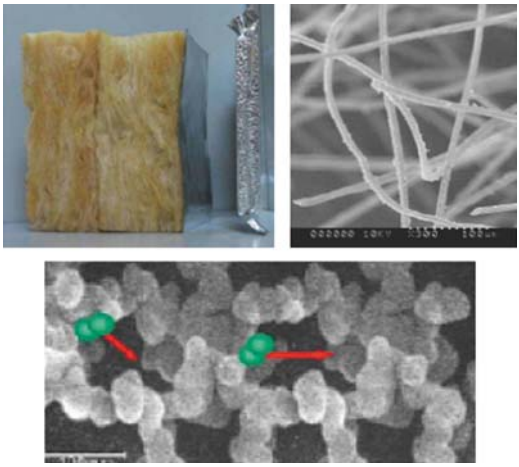

Fig. 1.2 *Top left*: Thickness of rock wool need to achieve the thermal insulation produced by an aerogel panel. *Top right*: Aerogel fibre with porosity greater than 90%. *Bottom*: Silica nanoparticle structure (CSTB)

[1] Apart from savings of raw materials, miniaturisation can increase operating speeds because electrons travel shorter distances, and it can reduce the heat power dissipated for a given amount of data because interconnect resistances and operating currents in transistors decrease with size.

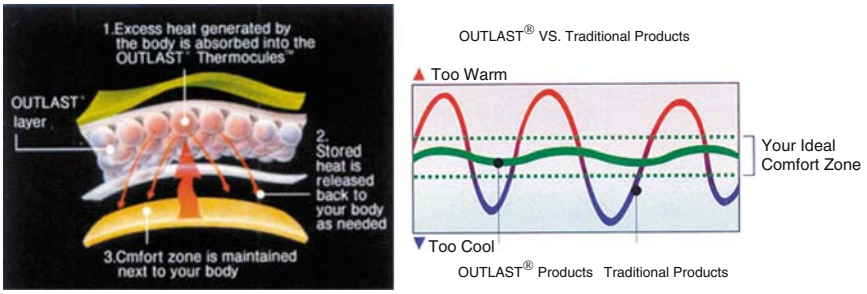
1.Excess heat generated by the body is absorbed into the OUTLAST® Thermocules™
OUTLAST® layer
2. Stored heat is released back to your body as needed
3.Cmfort zone is maintained next to your body
OUTLAST® VS. Traditional Products
Too Warm
Your Ideal Comfort Zone
Too Cool
OUTLAST® Products
Traditional Products

Fig. 1.3 *Left*: Structure of a cloth equipped with phase-change microcapsules called Thermocules. *Right*: Qualitative representation of temperature levels with (*green curve*) and without (*red and blue curve*) microcapsules

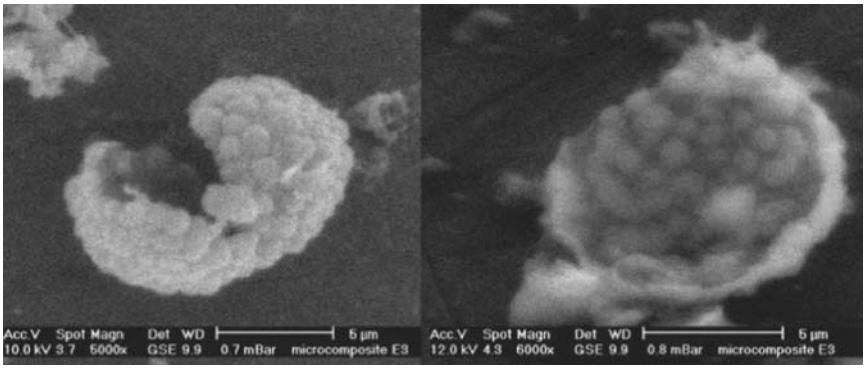
Acc.V Spot Magn Det WD 5 µm
10.0 kV 3.7 5000x GSE 9.9 0.7 mBar microcomposite E3
Acc.V Spot Magn Det WD 5 µm
12.0 kV 4.3 6000x GSE 9.9 0.8 mBar microcomposite E3

Fig. 1.4 Scanning electron microscope images of a packet of phase-change microcapsules. LGMT Roubaix

temperature, or with only slight variations, in response to changes in internal or external conditions. The phase change requires encapsulation (see Fig. 1.4), and the micron or submicron size of the capsules is what allows fast phase changes to occur.

A major aim for the semiconductor industry is the cooling of chips in microelectronics. Strategies are multiscale and varied. There are techniques operating on the scale of the chip itself. A good example is the system of fins of millimeter lengths made up of packets of nanotubes, as shown in Fig. 1.5 [4]. The very high thermal conductivity of the nanotubes makes the fins highly efficient. An increase of 19% has been demonstrated in the extracted power.

Undoubtedly the most exciting field of applications of nanostructures, and the one which has generated the most research and led to the most significant developments, is the area of thermoelectric conversion. Efficient materials for this must have high electrical conductivity and low thermal conductivity. Such paradoxical behaviour can be obtained by reducing the thermal conductivity by introducing nanostructures, e.g., nanoparticles (see Fig. 1.6) [5], nanowires (see Figs. 1.7 and 1.8) [6], or superlattices [7]. The merit factor was increased to 2.4 at room

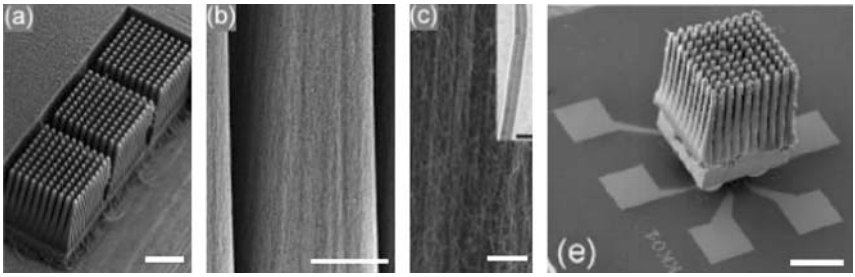

Fig. 1.5 Fins made from nanotubes for cooling electronic components [4]

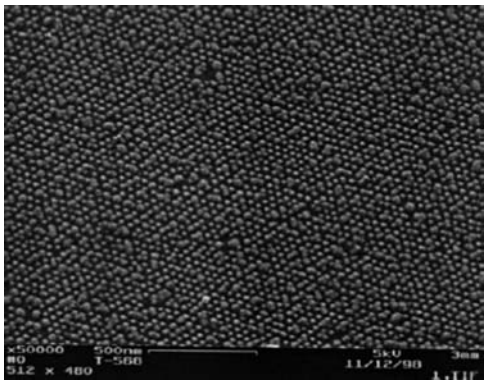

Fig. 1.6 PbSeTe quantum dots deposited by MBE. A merit factor of 1.6 has been achieved at room temperature [5]

temperature, for example, using superlattices, while it had remained below unity during the second half of the twentieth century.

The emerging area of thermal diode nanostructures was largely triggered by recent work by A. Majumdar and coworkers at UC Berkeley [8]. Increasing the mass of one end of a carbon nanotube (see Fig. 1.9) caused an asymmetry in the measured heat fluxes, as shown in Fig. 1.10, where the blue and red curves indicate the flux in the two directions in the tube. This asymmetry, although slight, has many physical implications, since classical thermodynamics and heat transfer forbid such behaviour. One can appeal to wave effects, but the waves commonly considered as heat carriers, viz., phonons, have completely symmetric behaviour when passing through an interface or a body. The paper [8] thus suggests the existence of rather special waves called solitons, likely to occur in a low-dimensional crystal lattice with a non-symmetric transmission factor, i.e., in which the transmitted amplitude depends on the direction of the wave through the crystal.

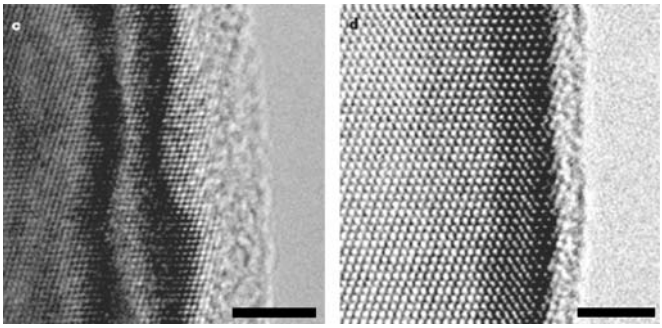

Fig. 1.7 Structural characterisation by scanning electron microscope of rough nanowires. *Left*: High-resolution transmission electron microscope image of a rough wire. The roughness appears between the wire and the native amorphous silica. *Right*: Transmission electron microscope image of an untreated nanowire. *Scale bars* represent 4 nm and 3 nm, respectively

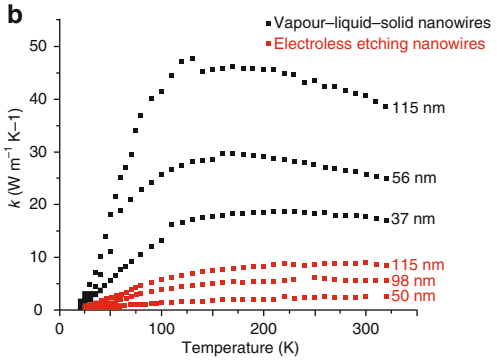

Fig. 1.8 Thermal conductivity of silicon nanowires with smooth surfaces (*black curves*) and chemically etched surfaces (*red curves*). A drop by a factor of 4 to 5 is observed when the surfaces are etched

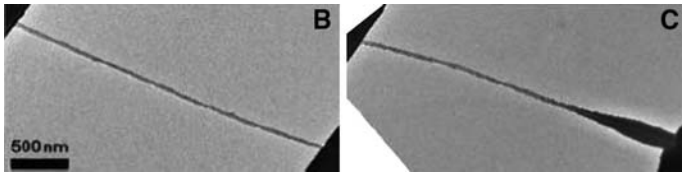

Fig. 1.9 Transmission electron microscope images of a suspended nanotube before (**b**) and after (**c**) adding C9H16Pt [8]

1.3 Physical Mechanisms

Heat conduction in nanostructures is not the same as in macroscopic systems, where it is characterised by Fourier's law. In the latter case, the heat carriers can be visualised as behaving like little beads with Brownian-like trajectories, i.e., suffering

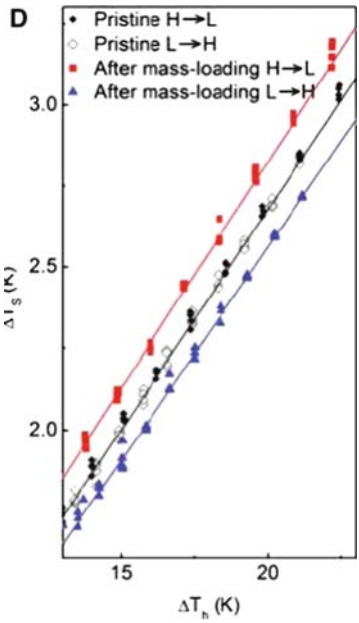

Fig. 1.10 Heat flux passing through a carbon nanotube for different imposed temperatures. When no mass is added to one end of the nanotube, measurements show that there is no difference between the two flow directions (*diamonds*). In contrast, the *blue* and *red curves* reveal this asymmetry when a mass is added to one end of the nanotube

frequent and random changes of direction. These movements are due to collisions between the carriers owing to their high density.

1.3.1 Rarefaction. Surface Reflection and Transmission at Interfaces

However, when this density decreases, the distance travelled by a heat particle between two collisions can exceed the characteristic length scale of the structure. The particle will then enter into more collisions with the walls of the system than with its counterparts within the system. This regime is no longer Brownian, but ballistic, because the particle will basically move in a straight line at constant speed between consecutive reflections from the system walls (see Fig. 1.11).

This is the first non-Fourier effect which could be qualified as a rarefaction phenomenon. The key mechanism here is the reflection of particles at the surface, but also transmission at the interface in the case of joined structures. If the reflection is perfectly specular, for example, the incident energy is fully redistributed in the symmetrical direction, and the flux component parallel to the wall remains unchanged, so there is no effect due to rarefaction in this same direction, even if the characteristic length scale is nanometric. But if now the reflection is diffuse and isotropic, i.e., all the energy is redistributed equally in all directions, a back flux arises and physical properties are modified.

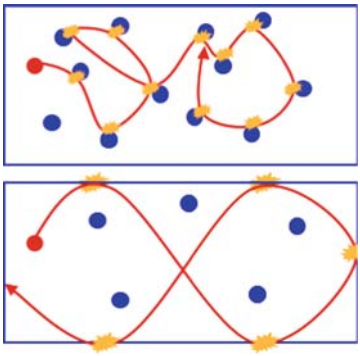

Fig. 1.11 Path of a heat carrying particle (*red blob*) in the diffusive regime (*top*), where the carrier density is high, and in the ballistic regime (*bottom*), where the carrier density is low. The prevailing mechanism in the diffusive case is interparticle interaction. In the ballistic case, it is particle–surface interactions that dominate

In a first approximation, reflection and transmission are assumed to be a linear combination of the two extremes, specular and diffuse. The fraction of the incident energy that is reflected specularly defines a coefficient called the specularity. But while the particle on its straight line path is indeed treated as a particle, it is its wavelike behaviour that governs its reflection or transmission. The specularity coefficient will thus depend on the wavelength and polarisation of the particle, the roughness of the surface, and the angle of incidence. It is therefore impossible to account exhaustively for the full complexity of the physical mechanisms that contribute to this coefficient, and it is generally treated as a floating parameter when computations are carried out.

This first rarefaction effect is often computed using the Boltzmann equation, which expresses the conservation of the number of heat carrying particles. Numerical solution can be based upon a classical approach or a direct method such as Monte Carlo simulation. The weak point in such simulations is the lack of data concerning the mean free path of the calculated mode, but also concerning the specularity coefficient. These methods will be examined in the next chapter.

1.3.2 Confinement

The word ‘particle’ is used to cover the more detailed reality of a localised wave packet. This wave packet is made up of several waves in different resonant or normal modes. It is the mode, the manner of vibration, that contains the energy of the system. It is assumed to be a travelling wave, since the system is a bulk system and much bigger than the lattice constant, i.e., the interatomic distance. The amplitudes u_n of these waves can be modelled by plane monochromatic waves, that is, complex exponentials whose arguments contain the wave vectors k and a time dependence

associated with the frequency ω:

$$u_n = u\mathrm{e}^{\mathrm{i}(kx-\omega t)} .$$

When modelling such modes, the boundary conditions are called Born–Von Karman conditions: the wave arriving at one end will come back in by the other. It thus propagates indefinitely in the same direction as long as it does not interact, and it moves at a speed imposed by the speed of the given mode.

Imagine now that the wave amplitude is annihilated at one end. This is what happens, for example, at the bridge of a guitar or when an acoustic wave in a crystal arrives at a free surface. The wave incident at this stopping point will be reflected with reversal of its phase, as shown in Fig. 1.12. The incident and reflected waves can still be modelled by monochromatic plane waves, that is, complex exponentials whose arguments contain wave vectors k with opposite signs, since the waves propagate in opposite directions. Their superposition is thus modelled as a sum of two exponentials, equal to the product of a cosine function whose argument depends on the wave vector and a complex exponential defining the temporal phase:

$$u_n \sim \exp\mathrm{i}(kx-\omega t) + \exp\mathrm{i}(-kx-\omega t) = \cos(kx)\mathrm{e}^{-\mathrm{i}\omega t} .$$

The zeros of the cosine function do not depend on time and define the nodes of a stationary wave. The vanishing of the amplitude at the boundary $x = L$ requires $kL = \pi/2 + n2\pi$, where n is an integer.[2] The wavelengths are thus $L/(n+1/4)$, defining the normal modes of the cavity formed by the structure. If the width L varies, then the wavelengths will also vary. New eigenmodes specific to the nanostructure thus form.

This transformation of the travelling normal modes into stationary normal modes is the second non-Fourier effect, referred to as confinement. By definition, these stationary waves have zero propagation speed. As the heat flux is proportional to the speed, the contribution of such stationary modes to heat transfer also vanishes.

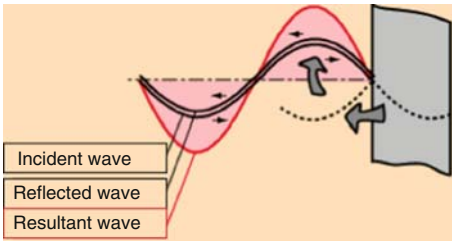

Fig. 1.12 An incident wave (*black line*) reflects (*black line*) on the surface with phase reversal. The superposition of the incident and reflected waves produces a stationary wave of twice the amplitude (*red line*)

[2] The root $n\pi$ not considered here would correspond to an incident wave moving away from the surface.

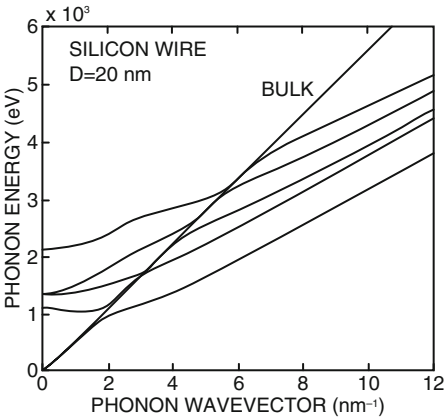

Fig. 1.13 Dispersion curves for the phonons in a silicon nanowire of diameter 20 nm (*continuous lines*). Small slopes and low group velocities correspond to small phonon wave vectors or long wavelengths as a result of confinement. Each branch is related to the projection of an oblique mode onto the wire axis. The bulk dispersion curve is shown by the *dashed line*

The dispersion curves giving the frequency as a function of the wave vector are therefore flat, because their slope is given by the mode speed, as shown in Fig. 1.13.

The second non-Fourier effect is thus determined by calculating these new eigenmodes. Analytical or numerical solutions of the elasticity equation can be implemented. They assume that the atomic motions can be treated as deformations of the crystal, itself treated as a continuum. This hypothesis remains doubtful for modes with short wavelengths. Approaches describing the motion of the atoms, such as lattice dynamics and molecular dynamics, remain more reliable but more difficult to apply at scales exceeding about ten nanometers.

1.3.3 Densities of States and Dimensionality

As can be seen from Fig. 1.13, confinement modifies the distribution of the modes as a function of frequency. The number of modes in a given frequency interval $[w, w+\mathrm{d}w]$ or wave vector interval $[\mathbf{k}, \mathbf{k}+\mathrm{d}^3\mathbf{k}]$ is called the density of states $D(\omega)$ or $D(k)$, respectively.

The directions of the vibrations in a crystal cover the whole space, and so do the directions of the wave vectors. The density of states in the bulk is thus proportional to a volume element, let us say an element in the form of a spherical shell, i.e., $D(k) \propto k^2 \mathrm{d}k$. If now the vibrations only build up in two dimensions, as in a graphene film, the density of states is proportional to a surface element, i.e., $D(k) \propto k\mathrm{d}k$. Finally, in the case of a nanowire, where the vibrations can only propagate in one direction, the density of states is proportional to a length element, i.e., $D(k) \propto \mathrm{d}k$.

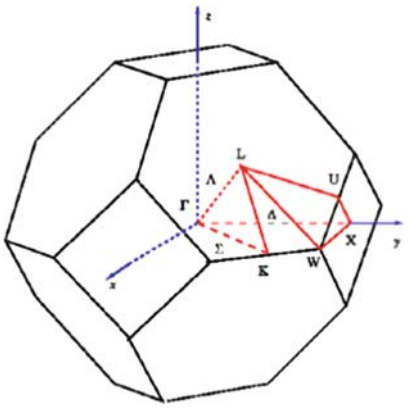

Fig. 1.14 Brillouin zone of an fcc crystal: the volume specified by the set of wave vectors of a bulk crystal. The space beyond this volume corresponds to wavelengths that are too short to be represented by the atoms of the crystal

In short, it can be shown that $D(k) = k^{d-1}\mathrm{d}k$, where d is the dimension of the structure. Since the thermal conductivity is proportional to the heat capacity, and hence to the density of states, the dimensionality of the structure has a significant impact on heat transfer. This is the third non-Fourier effect.

1.3.4 Non-Fourier Effects and Thermal Conductivity

In order to put the three non-Fourier effects into perspective, we shall establish a little known expression for the thermal conductivity [9] which contrasts the effect of the relaxation time, associated with the rarefaction phenomenon, and the effect of the density of states, reflecting confinement and the dimensionality of the structure.

We begin with an expression for the heat flux q:

$$q = \sum_k n_k \hbar \omega_k v_k \,, \tag{1.1}$$

taken as the product of the energy of mode k, i.e., the number n_k of particles in the mode multiplied by the energy $\hbar\omega_k$ of each such particle, and the group velocity v_k of mode k. This expression can be inserted into the Green–Kubo formula for the thermal conductivity [10], viz.,

$$\lambda = \frac{1}{V k_\mathrm{B} T^2} \int_0^\infty \big\langle q(0) q(t) \big\rangle \mathrm{d}t \,, \tag{1.2}$$

where V is the volume and k_B the Boltzmann constant, whence

$$\lambda = \frac{1}{Vk_B T^2}\int_0^\infty \mathrm{d}t \sum_k \left\langle n_k(0) n_k(t)\right\rangle (\hbar\omega_k v_k)^2$$
$$\propto \int_0^\infty \mathrm{d}t \int_k \mathrm{d}k (\hbar\omega_k v_k)^2 \left\langle n_k(0)^2\right\rangle \mathrm{e}^{-t/\tau_k} k^{d-1} . \quad (1.3)$$

The relaxation time has been denoted by τ_k. Here, cross-products between different modes have been dropped and the time dependence of the autocorrelation of the particle number has been expressed exactly [11]. The transition from a discrete sum to an integral[3] has brought in the density of states $k^{d-1}\mathrm{d}k$. To simplify, (1.3) is taken in the classical limit $n_k \hbar\omega_k = k_B T$. If the relaxation time $\tau_k = Ck^{-\delta}$, where δ can be viewed as the attenuation of mode k, then using the change of variable $u = tCk^\delta$, it follows that

$$\lambda \propto \int_0^\infty \mathrm{d}t \int_k \mathrm{d}k \, \mathrm{e}^{-t/\tau_k} k^{d-1} \propto \frac{1}{1-d/\delta}\left[t^{1-d/\delta}\right]_{t_{\min}}^{\infty} . \quad (1.4)$$

A minimal cutoff time $t_{\min}$, corresponding for example to the period of the fastest mode, has been introduced.

The Debye approximation is introduced in such a way that the group velocity can be treated as independent of the wave vector k and taken out as a constant factor. This is a rather crude approximation, because it ignores confinement, which should be taken into account through low group velocities that depend on the wave vector.[4]

If $d > \delta$, (1.4) becomes $t_{\min}^{1-d/\delta}/(1-d/\delta)$. This is shown in Fig. 1.15. The situation for the bulk material ($d = 3$, $d = 2$) and the possible case of a nanotube ($d = 2$, $\delta = 2$) are indicated. Figure 1.15 shows that, moving toward low values of d/δ, i.e., when the dimension of the structure decreases or the attenuation of the mode

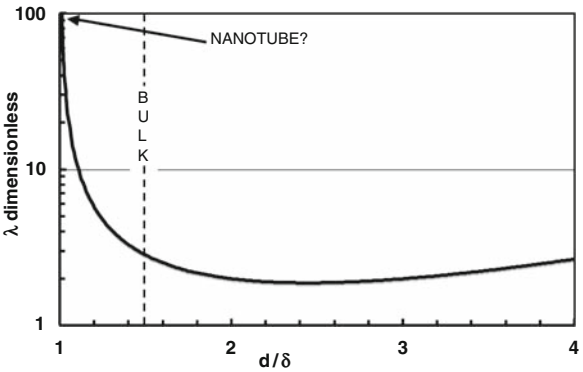

Fig. 1.15 Thermal conductivity as a function of the ratio d/δ, when $d > \delta$

[3] Note that the time integral of (1.3) leads to the well known formula $\lambda = \sum_k C_k v_k^2 \tau_k$ for the thermal conductivity, analogous to the result from kinetic theory.

[4] The increase in the number of branches due to confinement could be fairly easily accounted for by a sum over the branches of the result obtained, in which the constants C and δ would take different values in the different branches.

increases, the conductivity increases and even diverges. Moving toward high values of d/δ, the conductivity decreases to a minimum, then increases again. The two trends of increase and decrease can be put down to competition between attenuation and the reduction in the number of modes. If on the other hand $d = \delta$ or $d > \delta$, (1.4) clearly shows that the thermal conductivity becomes infinite.

1.4 Conclusion

An understanding of the way heat transfers in nanostructures will open the way to applications in the field of transport physics – typically using the Boltzmann equation and the description of rarefied regimes – and also in the field of solid-state physics – with the equation for atomic motions and phonon densities of states and dispersion curves. The various physical phenomena coming into play can lead to opposing trends for thermal properties. In a nanowire, for example, the thermal conductance is greatly reduced in comparison with its value in the bulk, while it is greatly increased in a single-wall nanotube.

In the next two chapters, the physics of the mechanisms introduced in this chapter is explored in more detail, and methods of solution are applied to establish quantitative properties of the basic nanostructures, viz., films, wires, and tubes.

References

1. S. Volz, G. Chen: Appl. Phys. Lett. **57**, 2056 (1999)
2. N. Mingo, D.A. Broido: Phys. Rev. Lett. **95**, 096105, 2005
3. G. Domingues, D. Rochais, S. Volz: J. Comp. Theor. NanoSc. **5**, 2, 153 (2008)
4. K. Kordás, G. Tóth, P. Moilanen, M. Kumpumäki, J. Vähäkangas, A. Uusimäki, R. Vajtai, P.M. Ajayan: Appl. Phys. Lett. **90**, 123105 (2007)
5. T.C. Harman, P.J. Taylor, M.P. Walsh, B.E. LaForge: Science **297**, No. 5590, 2229–32 (2002)
6. A.I. Hochbaum, R. Chen, R. Diaz Delgado, W. Liang, E.C. Garnett, M. Najarian, A. Majumdar, P. Yang: Nature **451**, No. 6381, 163–167 (2008)
7. R. Venkatasubramanian, E. Siivola, T. Colpitts and B. O'Quinn: Nature **413**, 597 (2001)
8. C.W. Chang, D. Okawa, A. Majumdar, A. Zettl: Science **314**, No. 5802, 1121–24 (2006)
9. S. Lepri, R. Livi, A. Politi: Physics Reports **377**, 1–80 (2003)
10. M. Toda, R. Kubo: *Statistical Physics II*, Springer Verlag, 2003
11. G.P. Srivastava: *The Physics of Phonons*, Taylor et Francis, 1990

Chapter 2
Nanostructures

Patrice Chantrenne, Karl Joulain, and David Lacroix

2.1 Introduction

As stated in Chap. 1, when the size of a solid object becomes of the same order of magnitude as the mean free path of the energy carriers, heat transfer is no longer diffusive. The notion of thermal conductivity, defined by Fourier's law for the diffusive regime, can then no longer be used. However, the thermal conductivity is such a common thermophysical parameter that this definition is still used when energy transport is non-diffusive. An equivalent thermal conductivity is then used which depends on the shape and size of the solid, the temperature, and the temperature gradient. The latter parameter is often not taken into account and this may be a source of error.

Nanostructures such as nanoparticles are of great interest in many applications. They are candidates for biomedical applications such as drug delivery and thermal treatment of cancer. They are used in nanofluids to improve convective heat transfer, with or without phase change (boiling, condensation). Nanotubes, nanowires, and nanofilms are widely considered in microelectronic applications as components, connecting wires [1], and sensors. Nanostructures are also of great help in physics for various experiments [2, 3]. Finally, all kinds of nanoparticles are used in nanocomposite materials. Most nanostructures are made of dielectric materials (mainly due to the importance of microelectronic applications), although some are made of electrically conducting materials.

In each application, the nanostructure interacts with its surroundings, and of course heat transfer in nanostructured materials or systems depends on these interactions, but also on their intrinsic thermal properties. When the thermal control of the system of interest is important, some knowledge of the thermal properties of each nanostructure is required. This knowledge may come from either experimental determination or theoretical prediction. But it is no easy matter to handle this aspect of nanostructures, and in many cases they have not been thermally

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_2, © Springer-Verlag Berlin Heidelberg 2009

characterised. Each nanostructure requires a specific experimental setup, which is often itself nanostructured.

A brief review of this experimental work is proposed in the appendix on p. 58. Compared with experimental investigation, many more publications are devoted to modelling heat transfer in nanostructures and to predicting their thermal conductivity. In these papers, previous experimental results are used to validate the theoretical approach. Section 2.2.1 reviews the physics of the vibrational properties of a semiconductor that underlie heat transport in such materials. This basis is essential for understanding the numerical models discussed in Sects. 2.2.2–2.3.2. Section 2.3.3 describes a technique that stands on its own, namely molecular dynamics, which also provides information on the thermal properties of nanostructures. In Sect. 2.4, simulation results obtained using these numerical models are presented for silicon and germanium nanostructures. Some numerical data for the application of molecular dynamics to carbon nanotubes is also discussed.

2.2 Modelling Heat Transfer

2.2.1 The Physics of Phonon Transfer

In this section, we discuss the theory of heat transport in solids, especially semiconductors. The books by Kittel [4] and Ashcroft and Mermin [5] are standard references here. The main points to be discussed are:

- the crystal lattice,
- dispersion relations,
- the notion of phonon and the quantisation of vibrational modes,
- the Boltzmann transport equation (BTE),
- the collisional relaxation time.

The Crystal Lattice

A crystal can be described as an assembly of identical elementary cells containing N_a atoms (see Fig. 2.1). There are three times N_a polarizations and as many dispersion curves for each wave vector direction. The sampling of wave vectors in the first Brillouin zone[1] depends on the geometry of the nanostructure (see Fig. 2.2). The crystal is defined by the number of cells N_{ai} in the direction of each primitive vector of the elementary cell $\mathbf{a}_i$ ($i = 1,2,3$). Wave vectors describing the crystal vibrations are defined in the reciprocal space. They are the sum of

[1] The first Brillouin zone of an atom is defined as the volume enclosed by the surface surrounding the atom that is everywhere equidistant from the atom and its nearest neighbours. It derives its significance from the description of wave propagation in a periodic medium, where it can be shown that solutions can be completely characterised by their appearance in this zone.

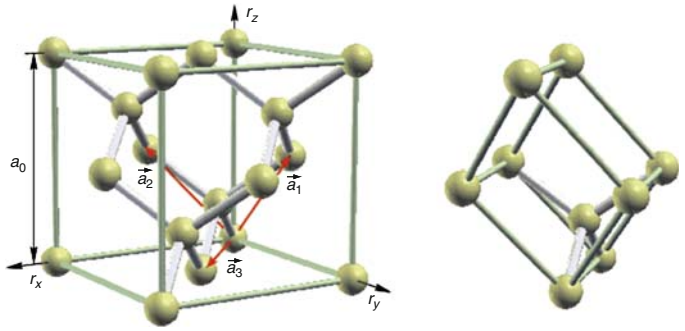

Fig. 2.1 (**a**) Silicon crystal lattice (diamond structure). Lattice parameter $a_0 = 5.431$ Å. The $\mathbf{a}_i$, $i = 1, 2, 3$, are the 3 primitive vectors. (**b**) Primitive cell of the silicon lattice

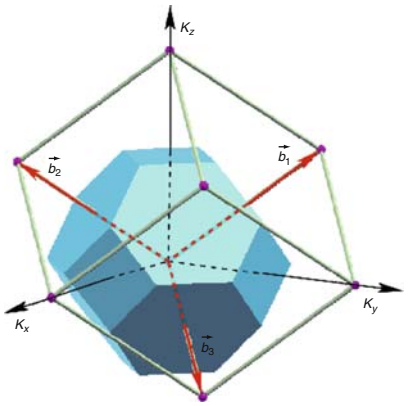

Fig. 2.2 Reciprocal lattice cell of silicon with first Brillouin zone, showing the reciprocal vectors $\mathbf{b}_i$, $i = 1, 2, 3$

$$\mathbf{K}_{bi} = \frac{n_{ai}}{N_{ai}} \mathbf{b}_i \,, \tag{2.1}$$

where $\mathbf{b}_i$ are the reciprocal vectors of $\mathbf{a}_i$. The limiting values of n_{a1}, n_{a2}, and n_{a3} are such that the wave vectors belong to the first Brillouin zone of the primitive cell.

Dispersion Relations

The dispersion relations express the relationship between the frequency ω and the wave vector $\mathbf{K}$ of a wave, or alternatively, between the energy $\hbar\omega$ and the momentum $\hbar\mathbf{K}$ of a particle. They thus characterise the vibrational properties of a medium. In the field of heat transfer, and more particularly, heat transport by conduction,

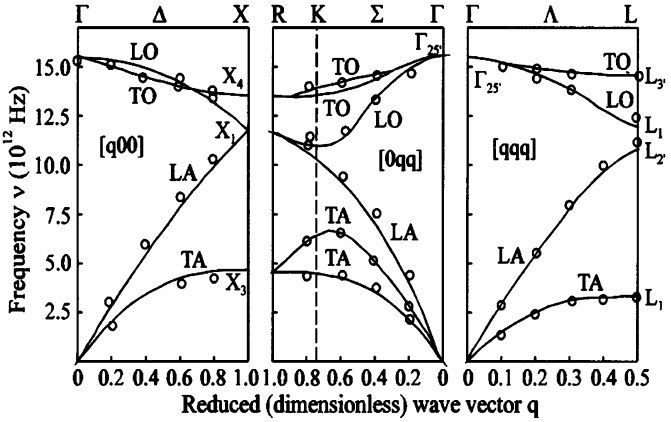

Fig. 2.3 Silicon dispersion relations. From R. Turbino et al. [6]

i.e., through vibrations in crystal structures, the dispersion relations of a material are used to define the velocity of the heat carriers, phonons or normal vibrational modes (eigenmodes) of the crystal. Three velocities are defined as a function of the frequency and polarisation of the wave: the group velocity v_g, the phase speed v_φ, and the speed of sound v_s. The speed of sound can also be defined as the low-frequency speed of acoustic modes, and the phase speed and group velocity tend to this quantity when ω is small:

$$v_g = \frac{d\omega}{dK}, \qquad v_\varphi = \frac{\omega}{K}, \qquad v_s = \lim_{K\to 0} \frac{d\omega}{dK}. \tag{2.2}$$

In the case of a very low temperature medium (a few tens of kelvin), the phonon speed can be assumed constant and taken equal to the speed of sound. One is then below the Debye temperature θ_D.

Experimentally, the dispersion relations are measured by inelastic neutron scattering in the planes [100], [011], and [111] of the bulk material. For silicon [6] (face centered cubic atomic structure), one obtains the curves shown in Fig. 2.3. Note the acoustic and optical branches for transverse and longitudinal polarisations (TA, LA, TO, and LO).

In most numerical simulations, one considers an isotropic medium in the reciprocal space. With this assumption, the dispersion relations can be approximated in various ways, e.g., a sine law with a single polarisation (Born–Von Karman) or a polynomial fit for the experimental curves. For an analytical determination of the thermal conduction properties, simple linear approximations are considered, over one or more frequency bands. For numerical solution of the Boltzman transport equation (BTE), we use the dispersion relations suggested by E. Pop et al. [7]. These have the advantage of being easy to differentiate, whence the group velocity and phase speed of the phonons can be calculated over the whole frequency spectrum (see Fig. 2.4).

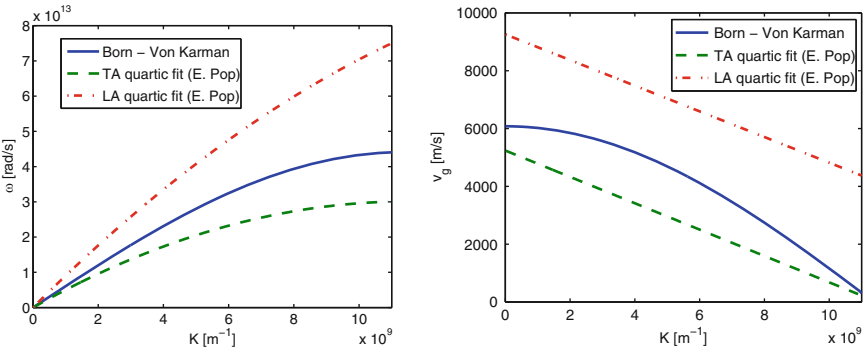

Fig. 2.4 Approximate spectral properties of silicon in the first Brillouin zone. *Left*: Dispersion curves. *Right*: Group velocities. Born–Von Karman model (*continuous curve*) [8] and polynomial fits (*dashed* and *dot-dashed curves*) [7]

Note. To calculate the heat transfer the optical branches of the dispersion relations are not needed. Indeed, the group velocity of the heat carriers from this branch is very low, and they contribute little to heat transport, whether it be ballistic or diffusive (see Chap. 1). However, for higher temperature applications, it may be necessary to consider these modes and the associated relaxation times. Han and Klemens [9] discuss this type of collisional interaction. On the other hand, when calculating the specific heat capacity of a material, these modes must be taken into account when the temperature is above the Debye limit ($T > \theta_{\mathrm{D}}$) [4].

Note. When using either experimental or approximate dispersion relations for thermal modelling of a nanostructure, one makes a crucial but tacit assumption, viz., that the dimensions of the structure remain large compared with the size of an atom (of the order of 1 Å). This bulk medium hypothesis is often justifiable. However, in the case of nanofilms or nanowires of the order of ten nanometers in thickness or diameter, such dispersion relations are no longer valid. The correct relations must be obtained by other means. One possible solution is to calculate them by considering the propagation of acoustic waves in thin wires or films [10]. They can also be determined by means of molecular dynamics and Green's functions [11]. But then another problem arises concerning the validity of the relaxation times, which are also defined via the bulk medium hypothesis.

Phonons, Quantisation, and Mode Density

Phonons are normal modes of vibration (or eigenmodes) of the crystal lattice. At thermodynamic equilibrium at temperature T, the quantum expectation of the phonon number $\langle n_{\mathbf{K},p} \rangle$ in some mode with angular frequency ω and polarisation p is given by the Planck (or Bose–Einstein) distribution.

$$\langle n_{\mathbf{K}} \rangle = \frac{1}{\exp\left(\hbar\omega/k_{\mathrm{B}}T\right) - 1} \,, \tag{2.3}$$

where k_{B} is the Boltzmann constant and $\hbar$ is Planck's constant divided by 2π. Each mode carries $\langle n_{\mathbf{K}} \rangle$ quanta of energy $\hbar\omega$. The average vibrational energy of a crystal is given by the sum over all wave vectors, including the various polarisations:

$$E = \sum_{\mathbf{K}} \langle n_{\mathbf{K}} \rangle \hbar\omega \,. \tag{2.4}$$

It is not easy to calculate this sum over all wave vectors $\mathbf{K}$ in the reciprocal space. It is usually preferable to integrate over the frequencies by introducing the so-called density of states $\mathfrak{D}(\omega)$. This gives the number of vibrational modes in the frequency range $[\omega, \omega + \mathrm{d}\omega]$. The 3D density of states of a crystal structure is calculated by counting the allowed values of $\mathbf{K}$ per unit volume $(2\pi/L)^3$ in the reciprocal (momentum) space. Assuming isotropic dispersion properties, for a crystal of volume $V = L^3$, where L is the side of a cube, we have

$$\mathfrak{D}(\omega)\mathrm{d}\omega = \frac{\mathrm{d}\mathbf{K}}{(2\pi/L)^3} = \frac{VK^2(\omega)\mathrm{d}K}{2\pi^2} \,, \qquad \mathfrak{D}_p(\omega) = \frac{VK^2(\omega)g_p}{2\pi^2 v_{\mathrm{g},p}(\omega)} \,, \tag{2.5}$$

where $v_{\mathrm{g},p}(\omega)$ is the group velocity for polarisation p. The factor g_p accounts for the degeneracy of the mode. For a transverse phonon, $g_{\mathrm{T}} = 2$, while for a longitudinal phonon, $g_{\mathrm{L}} = 1$. When considering a nanowire with two very small dimensions, the 1D mode density associated with suitable dispersion relations can also be used.

With this formalism, the energy E of a crystal can be rewritten as a function of the frequency ω and polarisation p of the phonons:

$$E = \sum_p \int_\omega \langle n_{\omega,p} \rangle \hbar\omega \mathfrak{D}_p(\omega)\mathrm{d}\omega \,, \tag{2.6a}$$

$$E = V \sum_p \int_\omega \left[\frac{\hbar\omega}{\exp\left(\hbar\omega/k_{\mathrm{B}}T\right) - 1} \right] \frac{K^2(\omega)g_p}{2\pi^2 v_{\mathrm{g},p}(\omega)} \mathrm{d}\omega \,. \tag{2.6b}$$

The energy expression (2.6b) will be used to obtain the local temperature (of a spatially discretised nanostructure) for numerical simulations solving the BTE.

Note. With the Debye approximation ($\omega = v_{\mathrm{g}}K$), the material is at low temperature and the group velocity constant. This energy can then be calculated formally. The derivative of E with respect to T yields a quasi-analytic form for the specific heat capacity C which is proportional to T^3. In the model provided by the kinetic theory, the thermal conductivity k can then be evaluated, with the result

$$k = \frac{1}{3} C v_{\mathrm{g}} \tau \,,$$

where τ is the collisional relaxation time.

The Boltzmann Transport Equation

The Boltzmann transport equation (BTE) describes the variation of the distribution function $f(t,\mathbf{r},\mathbf{K})$ (average phonon number at time t in the volume $\mathrm{d}^3\mathbf{r}$ around $\mathbf{r}$ with wave vector $\mathbf{K}$ given at $\mathrm{d}^3\mathbf{K}$):

$$\frac{\partial f}{\partial t} + \nabla_{\mathbf{K}}\omega\cdot\nabla_{\mathbf{r}} f + \mathbf{F}\cdot\nabla_{\mathbf{p}} f = \left.\frac{\partial f}{\partial t}\right|_{\mathrm{scat}} , \tag{2.7}$$

where the phonon number is

$$n(t,\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{K}} f(t,\mathbf{r},\mathbf{K}) , \tag{2.8}$$

$\partial f/\partial t$ is the time rate of change of the distribution function, $\nabla_{\mathbf{K}}\omega\cdot\nabla_{\mathbf{r}} f$ is the convective term (displacement of phonons within the structure), and $\mathbf{F}\cdot\nabla_{\mathbf{p}} f$ is the resultant of external actions on the phonon transport (neglected). The BTE is thus a classical transport equation associated with the phonon density. It can be solved using the mathematical tools developed for handling partial differential equations (see Sect. 2.3.1), or using statistical techniques (see Sect. 2.3.2).

Approximating the Scattering Term

There is still one term in this equation that is difficult to quantify, namely the scattering term $\partial f/\partial t\big|_{\mathrm{scat}}$. In the relaxation time approximation, it is assumed that the distribution function relaxes to the equilibrium distribution over some characteristic time τ:

$$\left.\frac{\partial f}{\partial t}\right|_{\mathrm{scat}} = -\frac{f_\omega - f_\omega^0}{\tau(\omega)} , \tag{2.9}$$

where f_ω^0 is the distribution function at thermodynamic equilibrium, given by the Bose–Einstein distribution. $\tau(\omega)$ is the collisional relaxation time used in the next section.

Radiative Formulation of the BTE

The BTE is the phonon transfer equation, and in this respect is very similar to the radiative transfer equation (RTE) for photons. The radiative exchanges have been modelled, and the RTE solved, for a wide range of media with various geometries, using a great many different techniques. It is thus useful to rewrite the BTE in a 'radiative' form, in order to take advantage of tools already developed in this field. The first step is to define the phonon spectral intensity $I_{\omega,p}$, i.e., the radiated intensity per unit area and frequency in an elementary solid angle $\mathrm{d}\Omega$ about the direction of propagation Ω of the phonon:

$$I_{\omega,p}(t,\mathbf{r},\Omega) = f_\omega(t,\mathbf{r})\mathfrak{D}_p(\omega)\hbar\omega v_{\mathrm{g},p}(\omega,\Omega)\ . \tag{2.10}$$

Bringing in the number of phonons $n_{\omega,p}$ in the frequency range, using (2.5) and (2.8), we have

$$n_{\omega,p}(t,\mathbf{r}) = f_\omega(t,\mathbf{r})\mathfrak{D}_p(\omega)\mathrm{d}\omega\ , \tag{2.11}$$

whence

$$I_{\omega,p}(t,\mathbf{r},\Omega) = n_{\omega,p}(t,\mathbf{r})v_{\mathrm{g},p}(\omega,\Omega)\hbar\omega\ . \tag{2.12}$$

The distribution function $f_{\omega,p}$ of (2.7) can be replaced by the phonon number $n_{\omega,p}$ to yield the phonon radiative transfer equation:

$$\frac{1}{v_{\mathrm{g},p}(\omega)}\frac{\partial I_{\omega,p}}{\partial t} + \Omega\cdot\nabla_{\mathbf{r}} I_{\omega,p} = -\frac{I_{\omega,p} - I^0_{\omega,p}}{\tau_p(\omega)v_{\mathrm{g},p}(\omega)}\ , \tag{2.13}$$

where $\tau_p(\omega)v_{\mathrm{g},p}(\omega)$ is inversely proportional to a spectral extinction coefficient $\kappa_p(\omega)$. The associated black body intensity becomes

$$I^0_{\omega,p} = \frac{\omega^2\mathrm{d}\omega}{2\pi^2 v^2_{\mathrm{ph},p}(\omega)v_{\mathrm{g},p}(\omega)}\frac{\hbar\omega}{\exp(\hbar\omega/k_{\mathrm{B}}T)-1}\frac{v_{\mathrm{g},p}(\omega)}{4\pi}\ , \tag{2.14}$$

the product of the density of states, the average energy per mode, and the velocity, uniformly distributed over 4π sr. (The intensity is a spectral and directional quantity.)

Collisional Relaxation Time

Scattering mechanisms are a key element when modelling heat transport in materials. Indeed, without collisions, the phonons would cross from one side of the structure to the other in accordance with the temperature gradient. The notion of thermal conductivity would then have little meaning, and one speaks of ballistic heat transport. This is what happens at very low temperatures, as we shall see later. In reality, phonons are more likely to interact in various ways, with the structure itself, but also with other phonons. These mechanisms fall into different categories and may be predominant or negligible depending on the temperature of the medium (see Fig. 2.5):

- Collisions with the boundaries of the structure. This is the predominant mechanism at very low temperatures, when the phonons have a very long mean free path. It also needs to be taken into account when the size of the structure is of the order of the phonon wavelength. Indeed, as the size of the nanostructure is reduced, phonon confinement is observed and with it a reduction in the thermal conductivity. This effect has been clearly demonstrated in silicon nanowires [12].
- Collisions with structural defects in the crystal lattice (holes, substitutions, dislocations, disclinations, and so on) or with atoms used for doping in the case of

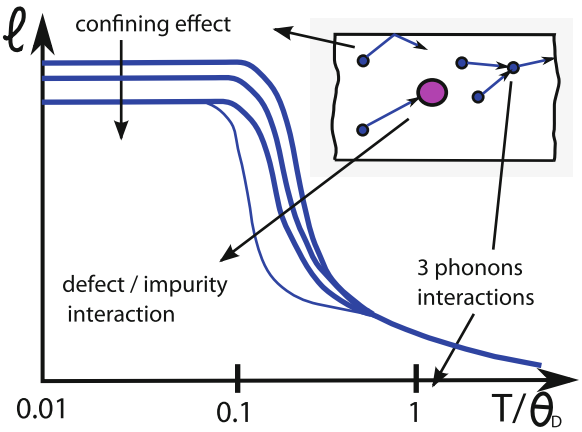

Fig. 2.5 Temperature dependence of the mean free path $\ell = v_g \tau$ in a material

semiconductors. These collisions dominate at intermediate temperatures and are proportional to the number of defects or dopants present in the crystal lattice.

- Collisions between phonons. The nature of such collisions depends on the polarisation and frequency of the phonons involved in the interaction. A distinction is made between what are called normal and umklapp processes. They dominate at room temperature and above.

During a collision, the phonon can change its direction of propagation and also its frequency, in accordance with the conservation of energy and momentum.

Collisions with Boundaries

The relaxation time τ_B associated with this type of collision depends solely on the characteristic size of the structure $\mathscr{L}$, the group velocity of the phonon, and in some cases a fitting factor F $(0 < F < 1)$:

$$\tau_B^{-1} \approx \frac{v_g F}{\mathscr{L}} \,. \tag{2.15}$$

This collision time is introduced into most analytical models for the simulation of thermal conductivity in order to obtain consistent values at low temperatures.

Collisions with Defects and Dopants

The collision time τ_I associated with point defects in the crystal lattice results from a modification of the vibrational properties due to the local change in mass Δm. According to P.G. Klemens [13], this relaxation time is proportional to ω^4 for a

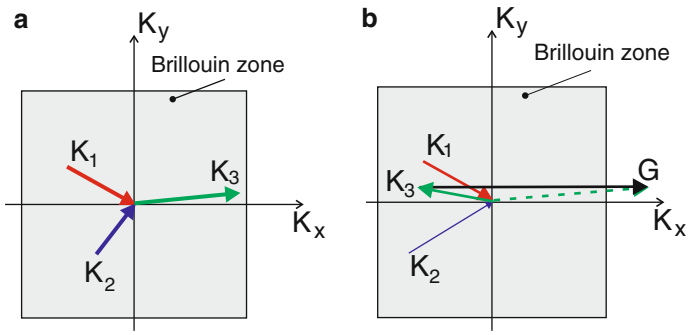

Fig. 2.6 Normal (*left*) and umklapp (*right*) processes. Two-dimensional representation of the reciprocal lattice

point defect (Rayleigh scattering):

$$\tau_{\mathrm{I}} \approx \frac{V}{4\pi v_{\mathrm{g}}^3} \left(\frac{\Delta m}{m} \right)^2 \omega^4 \,. \tag{2.16}$$

Collisions Between Phonons

Collisions between phonons are anharmonic processes involving three phonons (or sometimes four, at high temperatures). During such a collision, the phonons must satisfy conservation of energy and momentum, which leads to the following relations between frequencies and wave vectors:

$$\omega_1 + \omega_2 = \omega_3 \,, \tag{2.17a}$$

$$\mathbf{K}_1 + \mathbf{K}_2 = \mathbf{K}_3 \qquad \text{(normal process)} \,, \tag{2.17b}$$

$$\mathbf{K}_1 + \mathbf{K}_2 = \mathbf{K}_3 + \mathbf{G} \qquad \text{(umklapp process)} \,. \tag{2.17c}$$

The two possible mechanisms, normal and umklapp, result from the periodic arrangement of the crystal lattice (see Fig. 2.6). If the sum of the two wave vectors yields a vector $\mathbf{K}_3$ within the first Brillouin zone, one speaks of a normal process. However, if the resulting vector has amplitude above $\mathbf{K}_{\max}$, the latter is brought back within the first Brillouin zone by adding a vector $\mathbf{G}$ from the reciprocal lattice. This is an umklapp (reversal) process. Umklapp processes underlie the resistive mechanism associated with heat diffusion (Fourier law).

The collision times τ_{N} and τ_{U} for these two processes depend on the frequency and polarisation of the phonons, and also the temperature. Investigations by Klemens, Callaway, and Holland have greatly contributed to identification of these quantities [14–16]. Figure 2.7 provides a simplified representation of the possible interactions between transverse and longitudinal phonons.

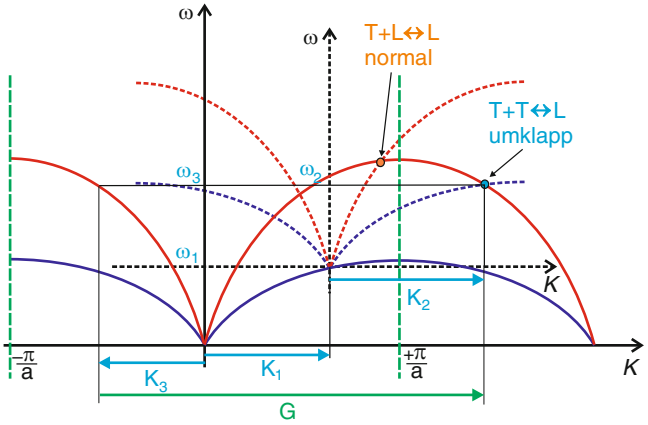

Fig. 2.7 Possible interactions between phonons with different polarisations

Matthiessen's Rule

All the collisional relaxation times discussed above are considered to be independent. In this case, an overall relaxation time τ_{tot} can be defined. It is given by the so-called Matthiessen rule:

$$\frac{1}{\tau_{tot}} = \frac{1}{\tau_B} + \frac{1}{\tau_I} + \frac{1}{\tau_N} + \frac{1}{\tau_U} \,. \tag{2.18}$$

Note. When the discrete ordinate or Monte Carlo methods are used, the relaxation time τ_B (interaction with boundaries) is not explicitly included insofar as phonon collisions with the walls are exactly taken into account during the displacement of the phonons within the structure.

2.2.2 Semi-Analytic Models

Linear Approximation to Boltzmann

In 1951, Klemens [14] was certainly the first to put forward a general formula for calculating the thermal conductivity of dielectric crystals. This formula is derived from the solution to the linearised Boltzmann equation for the phonon distribution function in the relaxation time approximation. J. Callaway [15] and M.G. Holland [16] used this formulation to predict the thermal conductivity of semiconductor materials, but taking the various phonon diffusion phenomena and vibrational behaviour of the materials into account in different ways.

For nanostructures, two phenomena affect the thermal conductivity:

- Phonon confinement. In order to account for this effect, A. Balandin and K. Wang [10] went back to Callaway's formulation, integrating the dispersion curves of a wire obtained by the theory of elasticity to calculate the relaxation time. Still for nanowires, N. Mingo [11] determined the dispersion curves from the interatomic interaction potential.
- Modification of the phonon distribution function due to the presence of free surfaces. This modification depends on the proportion of phonons that are specularly reflected relative to those that are diffusely reflected.

Kinetic Theory of Gases

This model was developed to predict the thermal conductivity of nanostructures made of dielectric crystals. If the distribution of the phonon propagation directions is uniform and phonon properties are isotropic, then the kinetic theory (KT) of gases may be applied to predict their thermal conductivity. Under these conditions, the thermal conductivity k_z in the z direction associated with phonons of wave vector $\mathbf{K}$ and polarization p can be written in the form [4, 17, 18]

$$k_{z,p}(\mathbf{K}) = C_p(\mathbf{K}) v_{\mathrm{g},p}^2(\mathbf{K}) \tau_p(\mathbf{K}) \cos^2\theta_z(\mathbf{K}) \,, \tag{2.19}$$

where v_g is the group velocity determined from the dispersion curves $v_{\mathrm{g},p}(\mathbf{K}) = \mathrm{d}\omega_p(\mathbf{K})/\mathrm{d}\mathbf{K}$, with ω the angular frequency, $\tau_p(\mathbf{K})$ is the mean phonon relaxation time due to phonon scattering phenomena, $\theta_z(\mathbf{K})$ is the angle between the wave vector $\mathbf{K}$ and the z direction, and $C_p(\mathbf{K})$ is the specific heat per unit volume for polarisation p, namely the temperature derivative of the internal energy $U_p(\mathbf{K})$. For a system of volume V [4],

$$C_p(\mathbf{K}) = k_\mathrm{B} x^2 \frac{\mathrm{e}^x}{V(\mathrm{e}^x - 1)^2} \,, \quad x = \frac{\hbar\omega}{k_\mathrm{B} T} \,. \tag{2.20}$$

The total thermal conductivity and specific heat are the sum of the individual contributions due to all the wave vectors $\mathbf{K}$ and polarisations p:

$$k_z = \sum_{\mathbf{K}} \sum_{p} C_p(\mathbf{K}) v_{\mathrm{g},p}^2(\mathbf{K}) \tau_p(\mathbf{K}) \cos^2\theta_z(\mathbf{K}) \,. \tag{2.21}$$

$$C = \sum_{\mathbf{K}} \sum_{p} C_p(\mathbf{K}) \,. \tag{2.22}$$

Under the assumption specified at the beginning of the section, the thermal conductivity should be isotropic. Actually, in our model, (2.21) is still used when the phonon properties are not isotropic. This anisotropy may be due to the dispersion curve or the relaxation time parameters.

For a given nanostructure made of a single crystal, the calculation of the thermal conductivity requires knowledge of the dispersion curves, the wave vectors, and the

relaxation time. This model is based on the assumption that the vibration properties of the nanostructure are the same as those of the bulk material. The bulk dispersion curves are used. As for the previous models, the relaxation time of a vibration mode is calculated using Matthiessen's rule from knowledge of relaxation times related to each scattering mechanism: normal phonon–phonon interactions, umklapp phonon–phonon interactions, interactions with defects, and interactions with the surfaces of the nanostructure. At the boundaries of the nanostructure, half of space is no longer accessible to the phonon, and the kinetic theory of gases should not then be used. However, the model was nevertheless constructed in this way, and the main features of the thermal conductivity of various nanostuctures are still picked up.

Ballistic Heat Transfer

Landauer's Formula

Consider a 1D system along the z axis, separating two thermostats at temperatures T_1 and T_2, as shown in Fig. 2.8. This system can be a film, nanowire, or nanotube. Each thermostat is treated as a phonon reservoir, i.e., all allowed phonon modes are those of an infinite volume. In particular, the density of states is that of an infinite medium.

Within the system, on the other hand, the modes are not the same as in the reservoir. A nanowire can behave like an acoustic waveguide in the same way as an optical fibre behaves as an optical waveguide. The modes are distributed over the polarisations p with dispersion relations of the form $\omega_p(K_z)$, where K_z is the component of the wave vector along the z axis. If these modes do not suffer collisions, it is a very straightforward matter to calculate the energy flux from reservoir 1 to reservoir 2 through the 1D system. The modes, populated at the temperature of one of the reservoirs, transport energy without interaction to the other reservoir through the nanostructure of length L. Regarding the energy flux, one then has the following expression, also known as the Landauer formula, which turns up in other transport problems in physics [19]:

$$q_z = \sum_{p,K_z} \hbar\omega_p(K_z) \frac{v_{g,p}(K_z)}{L} \left\{ \frac{\mathscr{T}_1(\omega)}{\exp\left[\hbar\omega_p(K_z)/k_B T_1\right] - 1} - \frac{\mathscr{T}_2(\omega)}{\exp\left[\hbar\omega_p(K_z)/k_B T_2\right] - 1} \right\}, \tag{2.23}$$

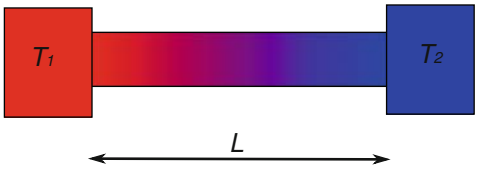

Fig. 2.8 Two thermostats at temperatures T_1 and T_2, a distance L apart

where $\mathscr{T}$ is the transmission coefficient from the reservoir toward the 1D system. There are various models for this coefficient, such as the diffuse acoustic mismatch model [20]. When the coefficient is equal to 1, the calculation is particularly simple. The sum over the wave numbers K_z is carried out by going to a continuous limit, introducing the 1D density of states

$$\mathfrak{D}(K_z)\mathrm{d}K_z = \frac{2\mathrm{d}K_z}{2\pi/L} = L\frac{\mathrm{d}\omega_p}{\pi v_\mathrm{g}} = \mathfrak{D}(\omega_p)\mathrm{d}\omega_p \,. \tag{2.24}$$

The energy flux q_z along the z axis then becomes

$$q_z = \sum_p \int_{\omega_{p,\min}}^{\omega_{p,\max}} \frac{\mathrm{d}\omega_p}{\pi} \left\{ \frac{\hbar\omega_p}{\exp\left[\hbar\omega_p(K_z)/k_\mathrm{B}T_1\right]-1} - \frac{\hbar\omega_p}{\exp\left[\hbar\omega_p(K_z)/k_\mathrm{B}T_2\right]-1} \right\} . \tag{2.25}$$

Carrying out the change of variable $x = \hbar\omega/k_\mathrm{B}T$, the flux becomes

$$q_z = \sum_p \frac{k_\mathrm{B}^2}{\pi\hbar} \left(T_1^2 \int_{x_{p,\min}}^{x_{p,\max}} \frac{x\mathrm{d}x}{\mathrm{e}^x-1} - T_2^2 \int_{x_{p,\min}}^{x_{p,\max}} \frac{x\mathrm{d}x}{\mathrm{e}^x-1} \right) . \tag{2.26}$$

Conductance Quantum

At sufficiently low temperatures, the sum over modes will only include modes passing through the origin. Likewise, the upper bound of the integral can be taken to infinity. Now $\int_0^\infty x\mathrm{d}x/(\mathrm{e}^x-1) = \pi^2/6$. When the temperatures of the reservoirs are close ($T_1 \sim T_2 \sim T$), the flux can be written in the form $q_z = NG_\mathrm{q}(T_1 - T_2)$, where G_q is the conductance quantum defined by

$$G_\mathrm{q} = \frac{\pi k_\mathrm{B}^2 T}{3\hbar} \,, \tag{2.27}$$

and where N is the number of modes passing through the origin. This conductance quantum represents the maximal contribution of each mode to the conductance of a 1D nanostructure. It does not really qualify as a quantum in the sense of the conductance quantum observed in electricity. It depends on the temperature. This phenomenon has been measured by Roukes and coworkers [21] for suspended structures.

Analytic Model of a Nanowire

With these considerations, the linearised Boltzmann equation in the relaxation time approximation gives rise to a different formulation for the thermal conductivity of a nanowire [22]:

$$k^\mathrm{wire}(T,\varepsilon) = k^\mathrm{bulk}(T,\varepsilon) - \Delta k^\mathrm{wire}(T,\varepsilon) \,, \tag{2.28}$$

with

$$k^{\text{bulk}} = \frac{k_\text{B}}{2\pi^2 v_\text{g}} \left(\frac{k_\text{B}}{\hbar}\right)^3 T^3 \int_0^{\theta_\text{D}/T} \frac{\tau x^4 \exp(x)}{\left[\exp(x)-1\right]^2} \mathrm{d}x \tag{2.29}$$

and

$$\Delta k^{\text{wire}}(T,\varepsilon) = \frac{24}{\pi} \left(\frac{k_\text{B}}{\hbar}\right)^3 \frac{k_\text{B}}{2\pi^2 v_\text{g}} T^3 \int_0^{\theta_\text{D}/T} \frac{\tau x^4 \exp(x)}{\left[\exp(x)-1\right]^2} G\big(\eta(x),\varepsilon\big) \mathrm{d}x\,, \tag{2.30}$$

where x is defined as before by $x = \hbar\omega/k_\text{B}T$.

The function $G\big(\eta(x),\varepsilon\big)$ depends on the ratio $\eta(x) = D/\ell(x)$ of the characteristic size D of the nanostructure and the mean free path of the phonons $\ell(x)$, as well as the surface roughness ε. J. Zou and A. Balandin [23] calculated this function for nanowires with circular cross-section, and X. Lu and J. Chu [24] for nanowires with square cross-section. Lu et al. [25] had already proposed a solution for nanowires with rectangular cross-section and had used the same formalism to predict the thermal conductivity of electrically conducting nanowires.

2.3 Nanofilms, Nanowires, and Nanotubes

2.3.1 Deterministic Model: BTE and the Discrete Ordinate Method

Introduction

As we have seen, the Boltzmann transport equation for phonons in the relaxation time approximation is perfectly analogous to the radiative transfer equation. Many tools have been developed to solve this equation. Among these, the discrete ordinate method (DOM) has proved particularly useful [26]. It was developed by Chandrasekhar [27] and is fully applicable to the case of phonon transfer in nanostructures. In this method, the intensity is calculated for a finite number of directions within a spatial lattice. Angle integrations used to obtain the heat flux or temperature are carried out using suitable quadratures. The application of this method to phonons is discussed in detail in [28, Chap. 5]. The discrete ordinate method is presented in both cylindrical and Cartesian coordinate systems.

DOM in Cylindrical Geometry

Cylindrical geometry is well suited to calculating the thermal behaviour of cylindrical nanowires, or that of nanofilms, in the direction normal to the film interfaces. Nanowires are systems in which the diameter D is nanometric and $D \ll L$, where L is the length of the wire. Films correspond to systems for which L is nanometric and D is infinite.

In the case of wires and films, the temperature is specified on the circular end faces of the cylinders. These faces are assumed to behave like phonon black bodies. The phonon intensity at the wall is equal to the intensity of a black body at the temperature of the wall. On the curved side wall of the cylinder, specular or diffuse reflection conditions are imposed. If the reflection conditions are purely specular, the phonon momentum will be conserved in the radial direction. The system is thus invariant in this direction and is equivalent to a film. When the reflection is diffuse (the incident intensity being redistributed in all directions), one has the situation in a nanowire. Reflection conditions with a diffuse part and a specular part are also possible in particular when the relevant phonon wavelengths are of the same order as the interface roughness.

In complete generality, the reflection conditions at a point $\mathbf{r}$ on the wall, in the direction Ω, can be written

$$I_{\omega,p}(\mathbf{r},\Omega) = \frac{\rho}{\pi} \sum_{\Omega'\cdot\mathbf{n}<0} I_{\omega,p}(\mathbf{r},\Omega')|\Omega'\cdot\mathbf{n}|\mathrm{d}\Omega + (1-\rho) I_{\omega,p}(\mathbf{r},\hat{\Omega}) \,, \tag{2.31}$$

where $\hat{\Omega}$ is the direction of specular incidence, Ω' are the other incident directions, and ρ is the ratio of diffuse to specular reflection. The case $\rho = 0$ thus corresponds to purely specular reflection.

In cylindrical geometry, the Boltzmann equation for the phonon intensity is found by rewriting (2.13) relative to the cylindrical coordinates (r,ϕ,z), whence

$$\frac{\mu}{r}\frac{\partial(rI_{\omega,p})}{\partial r} + \xi\frac{\partial I_{\omega,p}}{\partial z} - \frac{1}{r}\frac{\partial(\eta I_{\omega,p})}{\partial \phi} + \kappa_{\omega,p} I_{\omega,p} = \kappa_{\omega,p} I^0_{\omega}(T) \,, \tag{2.32}$$

where

$$\kappa_{\omega,p} = \frac{1}{V v_g \tau_{\omega,p}}$$

is the extinction coefficient, I^0_{ω} is the intensity at equilibrium, and μ, η, and ξ are the direction cosines specifying the direction Ω under consideration (see Fig. 2.9),

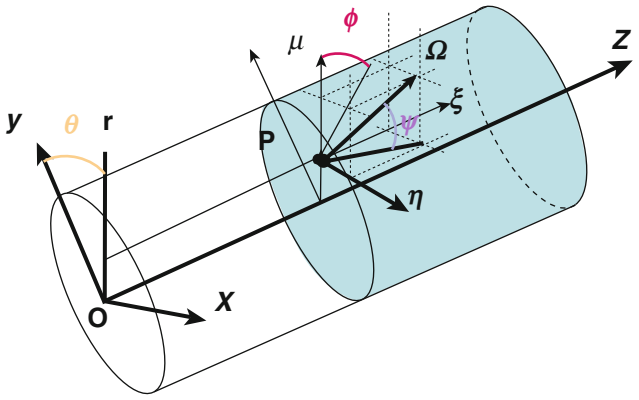

Fig. 2.9 Cylindrical geometry used to describe the phonon intensity

viz.,

$$\mu = \cos\phi \sin\psi \,, \qquad \eta = \sin\phi \sin\psi \,, \qquad \xi = \cos\psi \,, \qquad \mu^2 + \eta^2 + \xi^2 = 1 \,. \tag{2.33}$$

Thanks to the axial symmetry of this problem, the range of integration is limited to two dimensions in the plane (r,z), i.e., a longitudinal slice of the cylinder. At $r = 0$, a specular reflection condition must be imposed on the intensity in order to satisfy this symmetry. To obtain the intensity field, one can use the following iterative integration procedure. The equation is solved by starting from one of the surfaces on which the temperature is imposed, e.g., from right to left for directions pointing left. This solution is worked out given the initial temperature field in the medium. It is then solved in the same way starting from the second surface. The new temperature field is calculated at the end of the iteration by expressing the conservation of heat flux in the steady-state regime ($\nabla \cdot \mathbf{q} = 0$). The integral of the phonon intensity equation over the frequencies and solid angles is then

$$\int \kappa_\omega I_\omega \mathrm{d}\Omega \mathrm{d}\omega = \int \kappa_\omega I^0_\omega \mathrm{d}\Omega \,. \tag{2.34}$$

At each point the temperature is calculated so as to satisfy this relation. The iterative process is continued until the intensity field and temperature field have converged in accordance with a previously specified criterion. The resulting intensity field can be used to calculate not only the temperature field, but also the conductive heat flux due to the phonons within the structure. The latter can then be used to calculate the thermal properties of the structure, such as its conductance or its cross-plane conductivity.

This method of integrating the phonon intensity can of course be implemented in any coordinate system. For example, Cartesian coordinates can be used when we need to calculate the thermal properties of a nanofilm in the direction parallel to the interfaces (in-plane conductivity). In this case, the domain of integration is rectangular with height h and length L, where $h \ll L$. The temperature is imposed on the walls of height h, and diffuse reflection conditions are imposed on the walls of length L. The iteration process for obtaining the intensity field in the structure is the same as above.

2.3.2 Statistical Model: BTE and the Monte Carlo Method

Introduction

The statistical approach to phonon transport, also referred to as the Monte Carlo (MC) method for solving the BTE, is a technique using random processes to describe the displacement and collisions of packets of phonons. This type of approach, popular in the field of radiative transfer for its accuracy (photon transport via the RTE), is particularly well suited to solving the BTE in the relaxation time

approximation. With this method of solution, the transport aspect (phonon displacement) is decoupled from scattering mechanisms. Furthermore, as in most statistical techniques, the accuracy of the results is simply correlated with the number of quanta sampled. Reliable results can thus be obtained by this means, although the price to pay comes in computation time, which is also proportional to the sampling size.

As in the method described previously, the Monte Carlo method adapted to phonon transport characterises heat transfer in a nanostructure on the basis of energy considerations. Several aspects of the problem need to be discretised in order to formulate the problem:

- spatial discretisation of the structure,
- spectral discretisation of the vibrational properties of the material,
- temporal discretisation of the transport process.

This method also requires two types of input parameter in order to simulate heat transport. For one thing, the dispersion curves of the crystal structure of the material, and for another, an accurate evaluation of the collisional relaxation times for the relevant mechanisms, viz., normal and umklapp processes, scattering from impurities, etc. As with the discrete ordinate method, the Monte Carlo method provides access to temperatures and the heat flux within the nanostructure. Assuming a diffusive regime, one can thus estimate the thermal conduction. However, there is a further advantage in that complex 3D geometries can be represented in non-steady state regimes.

Bibliographical Notes

Statistical solution of the BTE has been undertaken on several occasions for semiconducting materials. The founding work by R.B. Peterson [29] assumes a 'Debye crystal' and thus gets around the problem of modelling the dispersion relations. Later, the model developed by S. Mazumder and A. Majumdar [30] improved Peterson's study by considering a realistic medium, viz., silicon. Our own contribution was to develop a technique in which energy is conserved during solution of the BTE [31]. We subsequently adapted this model to allow simulation of the thermal properties of silicon nanowires [32].

Statistical Solution of the BTE

Three Discretisation Processes

The three types of discretisation needed to solve the BTE via the Monte Carlo method are outlined here:

- Spatial discretisation. No particular geometrical constraint is required to discretise the spatial structure of a nanostructure when the Monte Carlo method is used to solve the BTE. Calculations are made by taking volume averages. The cells

covering the modelled region are therefore of comparable sizes in order to be able to justify these averages. Any kind of geometry is allowed, and objects with complex boundary configurations can thus be described. Moreover, no symmetry condition (either plane or axial) has to be respected in order to obtain simple solutions for the differential terms in the equation. For those structures we have investigated, viz., nanofilms and nanowires, a simple stack of cubic or cylindrical cells works perfectly.

- Spectral discretisation. The spectral discretisation of dispersion properties mentioned in Sect. 2.2.1 exploits the quadratic fit proposed by E. Pop [7] for silicon in the [100] plane. Given that the statistical solution of the BTE is spectral, the frequency range $0 < \omega < \omega_{\mathrm{LA\,max}}$ is divided up into equal intervals. $\omega_{\mathrm{LA\,max}}$ is the cutoff frequency for the longitudinal acoustic branch corresponding to $\mathbf{K}_{\max}$ (limit of the first Brillouin zone). One considers N_b spectral bands of width $\Delta\omega$, such that $\Delta\omega = \omega_{\mathrm{LA\,max}}/N_b$.
- Temporal discretisation. The choice of time step Δt is important in the simulation. It is determined partly by the length $\Delta l = v_{\mathrm{g}}\Delta t$ of the path to be travelled by the phonons, and partly by the scattering probability P_{scat}, taken to be given by

$$P_{\mathrm{scat}} = 1 - \exp\left(-\frac{\Delta t}{\tau_{\mathrm{tot}}}\right) . \tag{2.35}$$

The time step must thus be chosen small enough for two reasons. Firstly, in order to avoid ballistic transport of phonons through several cells of the discretised nanostructure, which would perturb the energy conservation principle. Secondly, if Δt is large compared with the collisional relaxation time, the scattering probability becomes equal to unity, and the interaction mechanisms between phonons are no longer correctly taken into account. The criterion generally applied is that $\Delta l < 10 l_{\mathrm{cell}}$, where l_{cell} is the length of one cell in the spatial lattice. The time step is then of picosecond order.

Sampling and Distribution Function

The second step in the calculation is to sample the phonon population in the modelled structure. Since the problem is not steady state, an initial condition must be defined. The usual initial data is the temperature field in the material. This field is generally assumed to be uniform, with heat exchanges resulting from the conditions applied on the boundaries. The latter are described below.

Sampling Frequency and Polarisation

A certain energy density corresponds to any given temperature T, as given by (2.6b), and each phonon of frequency ω transports the quantum $\hbar\omega$. One can thus deduce, at a given temperature T, the theoretical phonon number for the chosen spectral discretisation and for a unit cell of volume V:

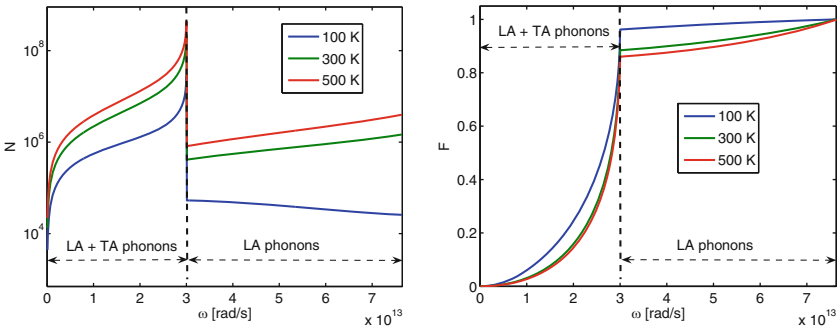

Fig. 2.10 Phonon sampling in a silicon unit cell of volume $V = 0.075$ µm^3 at three different temperatures: 100 K, 300 K, and 500 K. *Left*: Number of phonons. *Right*: Distribution function

$$N = V \sum_{p=\mathrm{TA,LA}} \sum_{b=0}^{N_b} \left[\frac{1}{\exp\left(\hbar\omega_{b,p}/k_\mathrm{B}T\right) - 1} \right] \frac{K_{b,p}^2}{2\pi^2 v_{\mathrm{g}\,b,p}} g_p \Delta\omega \,. \tag{2.36}$$

It is this quantity that is subsequently used to calculate the distribution function. The dependence on the density of states means that N does not vary monotonically given the polarisations considered (see Fig. 2.10 left). This in turn means that the frequency cannot be sampled directly. The standard technique for getting round this problem is to establish a cumulative distribution function that is then normalised. This function $F(\omega, T)$ is defined on the N_b spectral bands of the discretisation. For the frequency ω_i corresponding to the ith band, one sets

$$F_i(T) = \frac{\sum_{j=1}^{i} N_j(T)}{\sum_{j=1}^{N_b} N_j(T)} \,, \tag{2.37}$$

where $N_j(T)$ is the number of LA and TA phonons in the ith spectral band, defined by (2.36). The evolution of this distribution function is shown in Fig. 2.10 (right) for three different temperatures. Drawing a random number $\mathfrak{R}_i$ between 0 and 1 unambiguously yields a frequency. At 100 K, the phonons are mainly sampled at low frequency, since 96% of the phonons have frequencies below $\omega_{\mathrm{TA\,max}}$. At 500 K, this limit drops to 86%. The frequency resulting from the random draw lies in a band of width $\Delta\omega$ defined by the spectral discretisation. The sampled value ω_i is taken equal to

$$\omega_i = \omega_{0,i} + (2\mathfrak{R}_i - 1)\frac{\Delta\omega}{2} \,, \tag{2.38}$$

where $\omega_{0,i}$ is the frequency at the center of the ith band.

Once the frequency is known, the next step is to define the polarisation of the phonon. If the sampled frequency satisfies $\omega_i > \omega_{\mathrm{TA\,max}}$, the phonon lies on the LA branch. Otherwise, the polarisation must be determined randomly. The idea is to construct the probability of being on the longitudinal branch $P_\mathrm{LA}(\omega)$ from the

following ratio of the populations:

$$P_{\mathrm{LA}}(\omega_i) = \frac{N_{\mathrm{LA}}(\omega_i)}{N_{\mathrm{LA}}(\omega_i) + N_{\mathrm{TA}}(\omega_i)} \,. \tag{2.39}$$

A normalised random number $\mathfrak{R}_p$ is drawn. If $P_{\mathrm{LA}}(\omega_i) > \mathfrak{R}_p$, the phonon is longitudinal, otherwise it is transverse.

Sampling Direction and Position

The state of the particle is completely specified once the frequency and polarisation are known. In particular, one can deduce its wave vector $\mathbf{K}$ and group velocity v_{g}. The direction of propagation Ω of the phonon is fixed by sampling two random numbers $\mathfrak{R}_\theta$ and $\mathfrak{R}_\varphi$:

$$\Omega = \begin{cases} \sqrt{1-(2\mathfrak{R}_\theta - 1)^2}\cos(2\pi\mathfrak{R}_\varphi) \,, \\ \sqrt{1-(2\mathfrak{R}_\theta - 1)^2}\sin(2\pi\mathfrak{R}_\varphi) \,, \\ 2\mathfrak{R}_\theta - 1 \,. \end{cases} \tag{2.40}$$

The last step is to localise the phonon in the cell with dimensions $L_x \times L_y \times L_z$ located at $\mathbf{r}_{\mathrm{c}}$, in the Cartesian frame $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$. The position $\mathbf{r}$ of the phonon is given by three normalised random numbers $\mathfrak{R}_x$, $\mathfrak{R}_y$, and $\mathfrak{R}_z$:

$$\mathbf{r} = \mathbf{r}_{\mathrm{c}} + L_x\mathfrak{R}_x\,\mathbf{e}_x + L_y\mathfrak{R}_y\,\mathbf{e}_y + L_z\mathfrak{R}_z\,\mathbf{e}_z \,. \tag{2.41}$$

Weighting Factor and the End of Sampling

The initial sampling of the phonons in the statistical model is carried out in accordance with the local temperature of the cell, i.e., in accordance with its energy $E(T)$ as described by (2.6b). For temperatures above 100 K, it is clear that the number of phonons to be sampled to obtain the theoretical distribution given by (2.36) becomes quite considerable. As an example, in silicon, the number of phonons in a lattice cell of volume $V = 500\ \mathrm{nm} \times 500\ \mathrm{nm} \times 300\ \mathrm{nm} = 0.075\ \mu\mathrm{m}^3$ with population defined in Fig. 2.10 (left) and for all frequencies is given by

$$\begin{cases} T = 100\ \mathrm{K} \longmapsto N_{\mathrm{th}} = 6.23 \times 10^8\ \text{phonons} \,, \\ T = 300\ \mathrm{K} \longmapsto N_{\mathrm{th}} = 4.31 \times 10^9\ \text{phonons} \,, \\ T = 500\ \mathrm{K} \longmapsto N_{\mathrm{th}} = 8.79 \times 10^9\ \text{phonons} \,. \end{cases} \tag{2.42}$$

To model a structure of length $l = 6\ \mu\mathrm{m}$ at room temperature, around 90×10^9 phonons are needed! Even with the increased capacity of modern computers, this would not be feasible. To get round the problem, one can model 'packets'of phonons using a weighting factor W. For each frequency sampled, W phonons of the same frequency are attributed in the simulation. Sampling of the phonons stops when the energy obtained by summing the quanta $E^\star$ is equal to the energy $E(T)$ of the

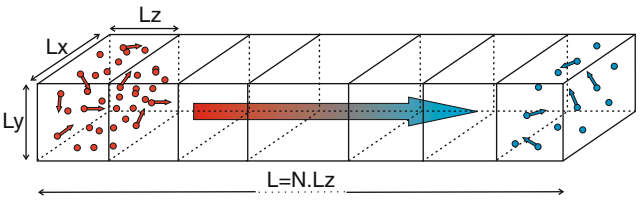

Fig. 2.11 Temperature imposed at the boundary of the nanostructure

crystal:

$$E^{\star} = \sum_{c=1}^{N_{\text{cell}}} \sum_{n=1}^{N^{\star}} W \hbar \omega_{n,c} \,, \tag{2.43}$$

where n is the index of the nth quantum sampled among $N^{\star}$ and c is the index of the cell in the lattice. The point about this weighting is to keep a statistically representative ensemble at different temperatures. If $T > 100$ K then $W > 1$, but if $T < 20$ K one sets $0 < W < 1$ so that realistic averages can be obtained.

Boundary Conditions

The boundary conditions on the surfaces of the calculation region will depend on the problem under consideration. To calculate the thermal conductivity of nanostructures, the cell temperatures are imposed at the ends of the region (see Fig. 2.11). These cells thereby become black bodies, wherein each phonon absorbed is reemitted at the imposed temperature. In a case where the diffusive regime is established, this type of boundary condition can be used to calculate the conductivity. It is also possible to impose a phonon flux through a wall.

In the case of adiabatic walls, phonons are reflected specularly. The energy and momentum of the heat carriers is then conserved. One must envisage the possibility of diffusely reflecting walls whenever the characteristic size of the nanostructure is of the order of the mean free path of the phonons. This is notably the case for nanowires with very small diameters. One can also consider a partly diffuse reflection and a partly specular reflection depending on the roughness of the surface of the nanostructure and the wavelength of the phonon incident upon it.

Transport and Collisions

Once the above initialisation stages have been validated, the iterative calculation can begin.

Phonon Displacements

For each phonon, a new position is determined within the discretised nanostructure depending on the time step Δt, the direction of propagation, and the group velocity.

If reflection from a wall occurs, a new orientation is attributed, depending on the nature of the wall. When all the phonons have been displaced, the energy $E^\star$ is recalculated within each cell by summing the quanta. Inverting (2.6), one obtains the local temperature $T^\star$ of each lattice cell in the nanostructure.

Scattering Interactions

Relaxation of the system to equilibrium involves collisions between phonons. They are defined in the context of the relaxation time approximation for the various processes, normal, umklapp, and collisions with impurities. The total relaxation time is given by (2.18). The scattering probability is defined by (2.35). A random number $\mathfrak{R}_{\text{scat}}$ is drawn, and there is a collision if $\mathfrak{R}_{\text{scat}} < P_{\text{scat}}$. Depending on the type of collision, the state of the phonon may be completely redefined (umklapp processes), or only partly (in normal processes, the direction of propagation is conserved, and for scattering from impurities, a new direction is sampled).

In the case of normal or umklapp collisions, the resulting phonon has a different frequency. In the framework of the Monte Carlo model discussed here, the total energy and momentum of all phonons present at a given instant in a cell is conserved. This involves modifying the cumulative distribution function associated with random selection of the parameters of scattered phonons, and this at the new cell temperature. One modifies the expression for the cumulative distribution function given by (2.37):

$$F_{\text{scat}}(T^\star) = \frac{\sum_{j=1}^{i} N_j(T^\star) P_{\text{scat}\,j}}{\sum_{j=1}^{N_b} N_j(T^\star) P_{\text{scat}\,j}} \,. \tag{2.44}$$

The resulting phonon sampling does indeed conform to the Planck equilibrium distribution. The sampling and the theoretical phonon distribution are compared at temperatures $T = 300$ K and $T = 500$ K in Fig. 2.12.

For a medium where the temperature is imposed on the boundary, the first and last cells are thermalised. At each time step, the phonons are resampled according to the initial distribution F. Under these conditions, collisions within these two cells are not calculated.

Iteration and the Resulting Variables

The phonon displacement and scattering processes are pursued until the steady-state regime is reached, which depends partly on the size of the nanostructure and partly on the chosen time step. In order to refine the results (temperature and flux profiles), these calculations are averaged over several identical simulations with different random number seeds.

The output variables are the temperature field $T(x,y,z)$ and energy field $E(x,y,z)$ in the discretised nanostructure, and also the heat fluxes $\mathbf{q}$ along the principal axes. The latter are obtained simply by summing the product $\hbar\omega\mathbf{v}_{\text{g}}$, where $\mathbf{v}_{\text{g}}$ is the velocity vector in a given direction. For the direction $\mathbf{e}_z$, one finds

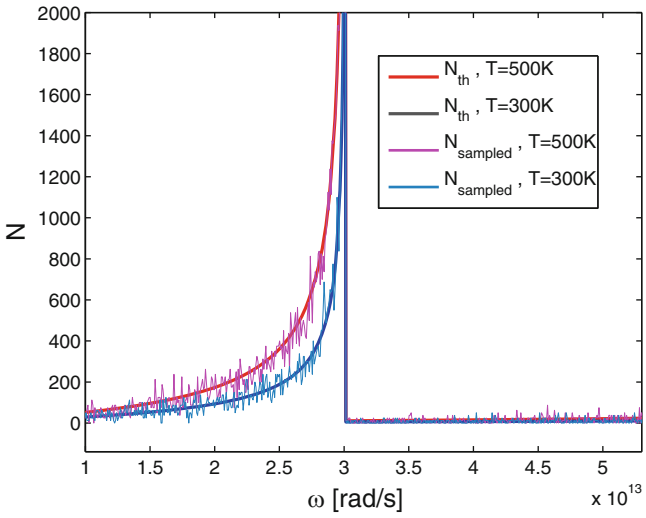

Fig. 2.12 Phonon frequency spectrum at 300 K and 500 K. Sampled and theoretical distribution

$$q_z = \sum_{n=1}^{N^\star} W \hbar \omega_n \mathbf{v}_{\mathrm{g},n} \cdot \mathbf{e}_z \, . \tag{2.45}$$

But the Monte Carlo model described here can be used to find other quantities, such as collisional interaction times with the structure. Using large populations of phonons, τ_B can be estimated statistically by observing the number of collisions at a given frequency. Simulation results obtained with this technique are discussed in Sect. 2.4.

2.3.3 Mechanical Model: Molecular Dynamics

As the name suggests, molecular dynamics is a numerical simulation technique that can be used to calculate the dynamics of an ensemble of atoms or molecules. The basic idea is to treat each atom as a point mass m with a position $\mathbf{r}$, a velocity $\mathbf{v}$, and an acceleration $\mathbf{a}$. At a given time, the acceleration is calculated from the total forces using Newton's second law:

$$\mathbf{F} = m\mathbf{a} \, . \tag{2.46}$$

The velocity and position fields are then found by integrating this equation with respect to time for all the atoms. The force $\mathbf{F}$ is the sum of the forces exerted by neighbouring atoms, derived from an interatomic potential [33–39], and forces due to external potential fields such as gravity, electromagnetic fields, and so on. The

molecular dynamics technique is described in detail in [28, 40–44]. A summary of its use to predict the thermal conductivity is given in [28, 45]. In this section, we discuss the limitations of molecular dynamics and the main methods used to predict the thermal conductivity.

Limitations of Molecular Dynamics for Studying Heat Transfer

The time step for integrating (2.46) is usually of femtosecond order, and the average volume occupied by an atom is less than the nm^3. As a consequence, computation times do not at present allow one to consider simulation times of more than a few tens or a few hundred nanoseconds, and the number of atoms is limited to a few million. In principle, molecular dynamics looks well suited to the study of nanostructures such as nanoparticles, nanowires, nanotubes, and nanofilms. However, the calculation time depends heavily on the complexity of the interaction potentials that are used. For materials like silicon, germanium, and carbon, the ones most often considered in applications, these potentials are particularly costly in this sense, and this limits the number of atoms. At the present time, the maximal dimensions that have been reached with molecular dynamics are just in the range of the very smallest nanostructures that can be made. For this reason, relatively few publications have been devoted to nanostructures. These concern carbon nanotubes with lengths from a few nm up to the μm [46–51], argon nanofilms [52], and more recently silicon nanofilms with thicknesses up to a few hundred nm [53]. The thermal conductivity of nanoparticles with characteristic dimensions of the order of a few tens of nm has been predicted for argon [46] and nickel (but only for energy transport by phonons) [54].

In molecular dynamics, the behaviour of atoms and molecules is described by classical mechanics. The main consequence of this description concerns the energy distribution over the various vibration modes of the system. All the vibration modes have the same energy $k_\mathrm{B}T$. The energy of a vibration mode (see Sect. 2.4.2) is equal to the quantum of energy of a phonon at the relevant frequency multiplied by the average number of phonons at this frequency [the Planck distribution (2.3)]. Conversely, in molecular dynamics, the average number of phonons is given by

$$\langle n_\mathbf{K} \rangle = \frac{k_\mathrm{B}T}{\hbar\omega} \,. \tag{2.47}$$

This is the limiting value of the number of phonons given by the Planck distribution when the temperature tends to infinity. In conclusion, for a system at a given temperature T, molecular dynamics overestimates the energy of the system and the phonon population. This important point is illustrated by a chain of harmonic oscillators comprising N point masses m connected by springs of stiffness K_r, only able to move in the x direction and subject to periodic boundary conditions (see Fig. 2.13).

At rest, the masses lie a distance a apart and the potential energy of the system is zero (zero force between the masses). The vibration modes of this chain of

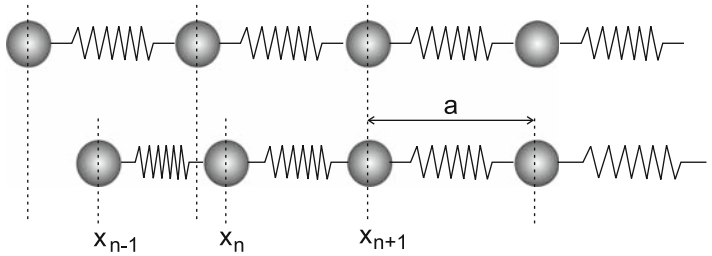

Fig. 2.13 One-dimensional harmonic system

oscillators are travelling plane waves characterised by the following dispersion relation:

$$\omega = 2\sqrt{\frac{K_r}{m}}\left|\sin\left(\frac{Ka}{2}\right)\right| , \quad \text{with } K = \frac{2\pi n}{aN} , \; n \in [1,N] . \tag{2.48}$$

The maximum angular velocity is $\omega_{max} = 2\sqrt{K_r/m}$. Figure 2.14 shows the temperature variations of the total energy of the system as calculated assuming a Planck distribution of the energy over the vibration modes, divided by the total energy of the system as calculated assuming a uniform distribution of the energy over the vibration modes. As expected, the ratio tends to the limiting value of 1 when the temperature increases.

In Fig. 2.15, the numbers of phonons in the classical and quantum systems are compared for two temperature values. At low temperatures ($k_B T/\hbar\omega_M = 0.32$), the classical system has a higher phonon number than the quantum system. At high temperatures ($k_B T/\hbar\omega_M = 80$), the phonon numbers are the same for the two systems.

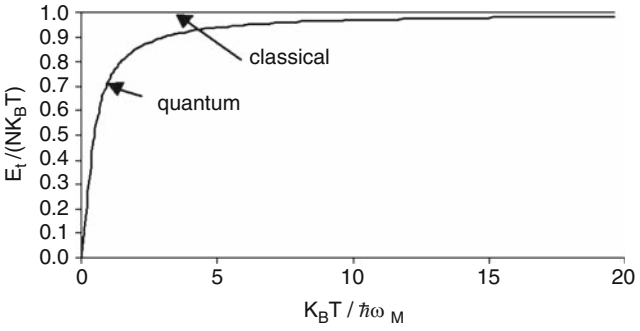

Fig. 2.14 Temperature dependence of the ratio of the total energy of a dimensionless quantum harmonic system to the total energy of the same system considered with a uniform distribution per vibrational mode

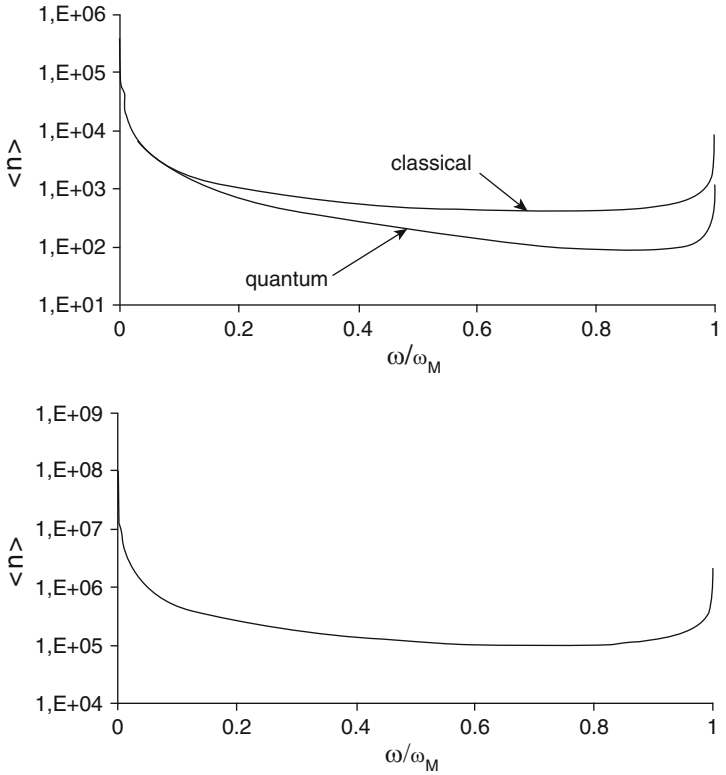

Fig. 2.15 Number of phonons as a function of the reduced (dimensionless) angular velocity. *Top*: $k_B T/\hbar\omega_M = 0.32$. *Bottom*: $k_B T/\hbar\omega_M = 80$. In the second case, the average phonon number calculated by the classical approach is equal to the one found by the quantum approach

Calculating the Thermal Conductivity

Three techniques are available to calculate the thermal conductivity using molecular dynamics simulations: equilibrium molecular dynamics (EMD), homogeneous non-equilibrium molecular dynamics (HNEMD), and non-homogeneous non-equilibrium molecular dynamics (NHNEMD).

Temperature calculations are a prerequisite for obtaining the thermal conductivity. Indeed, for a system at equilibrium, the thermal conductivity is calculated at a given temperature that must be known, and for a non-equilibrium system, the temperature gradient in the system must be determined. Whatever technique is actually used, the temperature calculation is based on the assumption of local thermodynamic equilibrium and equipartition of energy over all vibration modes and polarisations. For non-equilibrium systems, the local thermodynamic equilibrium hypothesis is based on the assumption of a small discrepancy between the actual distribution function and the Maxwell–Boltzmann velocity distribution function [55]. In a non-equilibrium system (solid argon at 25 K) with a temperature gradient of

3×10^8 K/m, it has been confirmed [17] that the difference between the local distribution functions and the Maxwell–Boltzmann distribution functions at the same temperature does not exceed 2%. In non-equilibrium molecular dynamics, owing to the small system sizes (a small multiple of the interatomic distance), temperature gradients are much higher than true temperature gradients. However, temperature levels always remain reasonable compared with temperature variations due to phase change. The instantaneous temperature of an ensemble of N atoms is given in terms of the average kinetic energy of these atoms:

$$\frac{3}{2}k_{\mathrm{B}}T(t) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}m_i v_i^2 \,. \tag{2.49}$$

The standard deviation of the temperature is proportional to the temperature and inversely proportional to the square root of the number of atoms [56]:

$$\sigma\big(T(t)\big) \propto \frac{T(t)}{\sqrt{N}} \,. \tag{2.50}$$

Consequently, if the number of atoms is large enough (of the order of 10^{23} for a macroscopic system), the statistical variations of the temperature are negligible. In molecular dynamics, the number of atoms considered varies from a few dozen to a few tens of thousands. The statistical variations in the instantaneous temperature are then no longer negligible. A time average must be calculated to reduce the standard deviation of the temperature:

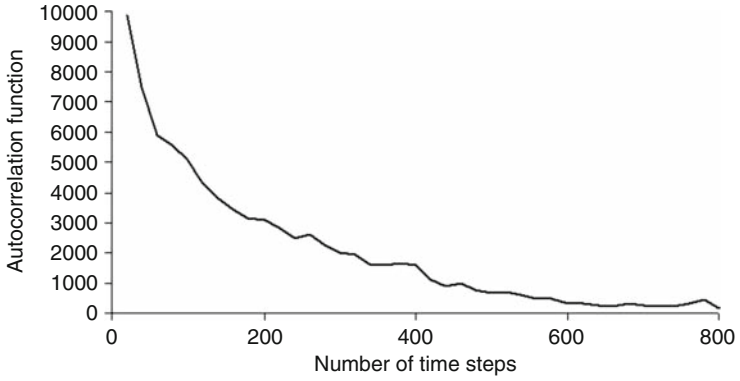

Fig. 2.16 Autocorrelation function of the local temperature in a system. Results obtained for a C(5,5) carbon nanotube of length 10.4 nm under a temperature gradient of 60 K at an average temperature of 900 K. Brenner interaction potential with time step 0.000 5 ps

$$\left\langle T(t)\right\rangle = \frac{1}{N_t}\sum_{i=1}^{N_t} T(t_i)\,, \tag{2.51a}$$

$$\sigma\Big(\left\langle T(t)\right\rangle\Big) \propto \frac{T(t)}{N_t}\sqrt{N}\,. \tag{2.51b}$$

Figure 2.16 shows the autocorrelation function of the instantaneous temperatures calculated at each time step during a molecular dynamics simulation. In this figure, the temperatures clearly remain strongly correlated over a period of a few hundred time steps. The time average is thus worked out using, for example, only one instantaneous temperature every 800 time steps. Apparently there is some debate about the correlation time. For example, Lukes et al. [52] define the autocorrelation time as being equal to the average phonon relaxation time.

Equilibrium Molecular Dynamics

The system considered is a microcanonical system at constant pressure, energy, and volume. As the system evolves without constraints, after a certain number of time steps, it reaches a thermodynamic equilibrium resulting in a constant average temperature. However, owing to temperature fluctuations, the instantaneous flux will also vary, even though its average value is zero since the system is at equilibrium. The conductivity calculation is based on the fluctuation–dissipation theorem of the linear response theory which relates transport properties to instantaneous fluctuations in the system. The thermal conductivity can thus be deduced from variations in the instantaneous flux density of a system at equilibrium using the Green–Kubo relation [40, 41, 57, 58]:

$$k = \frac{2V}{3k_{\mathrm{B}}T^2}\int_0^{\infty} \left\langle \mathbf{q}(0)\cdot\mathbf{q}(t)\right\rangle \mathrm{d}t\,. \tag{2.52}$$

In order to determine the thermal conductivity, one must therefore calculate the instantaneous flux density in the system [59]:

$$\mathbf{q}(t) = \sum_{i=1}^{N} \mathbf{v}_i E_i - \frac{1}{2}\sum_{i,j=1}^{N} \mathbf{r}_{ij}\mathbf{v}_i\mathbf{F}_{ij}\,. \tag{2.53}$$

The first and second terms represent respectively the kinetic energy and the potential energy transported by each atom moving with velocity $\mathbf{v}_i$. In solids, the first term is generally negligible [45].

The difficulty with this method lies in obtaining an accurate enough approximation for the integral in (2.52). This involves carrying out simulations over a rather large number of time steps, which may become prohibitive with regard to computation time when complex interaction potentials are used.

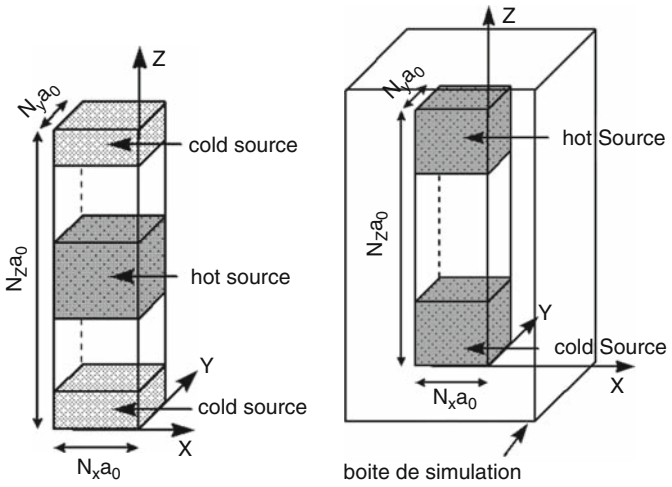

Fig. 2.17 Geometrical configurations for simulating heat transfer. *Left*: Periodic boundary conditions. The simulation box and system have the same size. *Right*: Free surface conditions. The simulation box is bigger than the system

Non-Homogeneous Non-Equilibrium Molecular Dynamics

The non-homogeneous non-equilibrium molecular dynamics technique is certainly the simplest to understand and implement, because it is equivalent to an experiment of the guarded hot plate type. It consists in simulating the 1D heat transfer in a system when a cold source and a hot source are placed in it, measuring the flux density exchanged between the sources and the temperature gradient between the sources. The definition of the cold and hot sources depends on the type of boundary conditions, viz., periodic or free surface (see Fig. 2.17). To simulate the heat transfer, one technique proposed in the literature consists in exchanging atoms between the hot and cold sources [60]. As this technique is not widely used, we shall not discuss it further. The most commonly used technique consists in modifying the velocity field of the atoms belonging to the heat sources in order to impose either the thermal power exchanged between the hot and cold sources [52, 61, 62], or the temperature of those sources [47, 63–66]. In theory, this modification of the velocity field must be carried out rather carefully to avoid artificially introducing a moment into the system. However, ZhenAn et al. [67] have shown that the moment introduced by such a modification of the velocity field by the techniques outlined below is negligible and does not introduce any significant error into the resulting value of the thermal conductivity.

Imposed Temperatures

The temperatures $T^*_{\text{hot}}(t)$ and $T^*_{\text{cold}}(t)$ of the hot and cold sources, respectively, are calculated after integrating Newton's equations. These temperatures differ from the

required temperatures T_{hot} and T_{cold}. The velocity field of the hot and cold sources is therefore multiplied by a weighting coefficient in order to correct these temperatures. The weighting coefficients are $\sqrt{T_{\text{hot}}(t)/T^*_{\text{hot}}}$ and $\sqrt{T_{\text{cold}}(t)/T^*_{\text{cold}}}$ for the hot and cold sources, respectively. With this method, the absolute values of the energies given up to the hot and cold sources are $\Phi_{\text{hot}}(t)$ and $\Phi_{\text{cold}}(t)$, respectively, given by

$$\Phi_{\text{hot}}(t) = \frac{3k_{\text{B}}}{2} N_{\text{hot}} \left[T_{\text{hot}} - T^*_{\text{hot}}(t) \right] , \tag{2.54a}$$

$$\Phi_{\text{cold}}(t) = \frac{3k_{\text{B}}}{2} N_{\text{cold}} \left[T_{\text{cold}} - T^*_{\text{cold}}(t) \right] , \tag{2.54b}$$

where N_{hot} and N_{cold} are the numbers of atoms in the hot and cold sources. With this method, the temperature gradient between the sources is perfectly controlled. The instantaneous flux densities exchanged by the heat sources are calculated from the energies $\Phi_{\text{hot}}(t)$ and $\Phi_{\text{cold}}(t)$:

$$q(t) = \frac{\Phi(t)}{S\Delta t} , \tag{2.55}$$

where S is the cross-sectional area traversed by the heat flux and Δt the time step.

In the steady-state regime, the average flux densities $\langle q_{\text{hot}}(t) \rangle$ and $\langle q_{\text{cold}}(t) \rangle$ must have the same absolute value. This is therefore a simple check for the principle of energy conservation. To illustrate the method, the temperature profile in a system is represented in Fig. 2.18 (upper). The system is a face centered cubic crystal of solid argon with dimensions $60a_0$ in the x, y, and z directions, where a_0 is the size of the cubic cell of the lattice. Periodic boundary conditions are used in the three directions. The hot and cold sources have thickness $12a_0$. Although the average temperatures of the hot and cold sources are held constant during the simulation, the temperature profile in these regions is not uniform. In fact it has a parabolic shape, typical of a macroscopic medium containing a heat source. Figure 2.18 (lower) shows the evolution of the average flux densities exchanged by the heat sources over 5 000 time steps. The steady-state regime is reached after about 50 ps.

Imposed Power

According to Lukes et al. [52], this method requires fewer time steps than the last in order to reach the steady-state regime, and this significantly reduces the computation time. Since the flux density q is imposed, the amount of energy that must be taken from the cold source and given to the hot source in each time step is given by (2.55). This energy (in absolute value) is supplied by modifying the kinetic energy of the hot source and the cold source. The velocity field is thus multiplied by a weighting coefficient given by

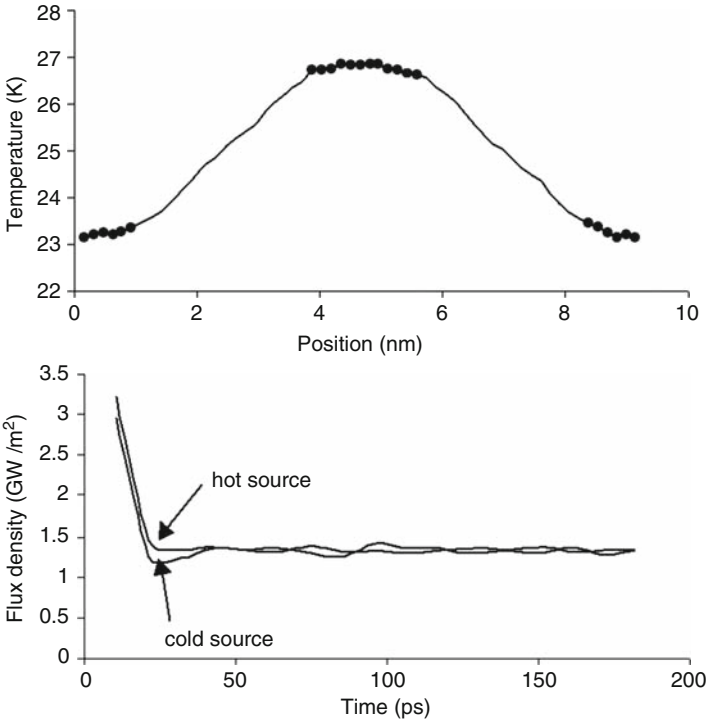

Fig. 2.18 Temperature and flux profiles in a system with thermostatically controlled hot and cold sources. *Left*: Temperature profile. *Dots* indicate the position of the sources. *Right*: Evolution of the average flux densities exchanged between the hot source and the cold source

$$\alpha_{\text{hot}}(t) = \sqrt{1 - \frac{\Phi}{E^{\text{kin}}_{\text{hot}}(t)}}\,, \tag{2.56a}$$

$$\alpha_{\text{cold}}(t) = \sqrt{1 - \frac{\Phi}{E^{\text{kin}}_{\text{cold}}(t)}}\,, \tag{2.56b}$$

for the hot and cold sources, where $E^{\text{kin}}_{\text{hot}}(t)$ and $E^{\text{kin}}_{\text{cold}}(t)$ are the instantaneous kinetic energies of the two sources.

The exchanged power is perfectly constant. The steady-state regime is reached when the average temperature profile is stable as time goes by. When the system does not conserve energy, a temperature drift appears. The average temperature of the system is generally close to the initial temperature.

Homogeneous Non-Equilibrium Molecular Dynamics

This technique was proposed by D.J. Evans [68, 69]. The idea is to apply a homogeneous external force field to the system in order to create a heat flux. At the same time, a Gauss thermostat [70, 71] is applied to hold the temperature constant. If $\mathbf{F}_e$ is the external force, the relevant equation of motion is

$$m_i \mathbf{a}_i = \frac{1}{2N}\left[\sum_j \mathbf{F}_{ij} + (E_i - E)\mathbf{F}_e - \sum_j \mathbf{F}_{ij}\mathbf{r}_{ij}\cdot\mathbf{F}_e + \sum_j \mathbf{F}_{jk}\mathbf{r}_{jk}\cdot\mathbf{F}_e\right] - \alpha m_i \mathbf{v}_i \,, \quad (2.57)$$

and the Gauss thermostat coefficient α is given by

$$\alpha = \sum_i (m_i \mathbf{v}_i \cdot \mathbf{v}_i) \sum_i \frac{m_i}{2N}\left[\sum_j \mathbf{F}_{ij} + (E_i - E)\mathbf{F}_e - \sum_j \mathbf{F}_{ij}\mathbf{r}_{ij}\cdot\mathbf{F}_e + \sum_j \mathbf{F}_{jk}\mathbf{r}_{jk}\cdot\mathbf{F}_e\right]\cdot\mathbf{v}_i \,. \quad (2.58)$$

Under these conditions, using perturbation theory, Evans et al. showed that the thermal conductivity is

$$k = \lim_{F_e \to 0} \frac{\langle q(t) \rangle}{F_e T} \,. \quad (2.59)$$

In order to determine the thermal conductivity, one must therefore calculate the average flux $\langle q(t) \rangle$ [the instantaneous flux calculated using (2.55)] for different values of the force $\mathbf{F}_e$, which must be small enough to qualify as a small perturbation as regards the equilibrium (linear perturbation approximation).

2.4 Comparison and Limitations of the Models

The different techniques discussed above have been used to model the thermal properties of nanostructured semiconductors and also to observe the heat exchanges occurring within nanostructures with various geometries and sizes. As mentioned in the introduction, the different models do not always provide access to the same information. The semi-analytic models, for example, give no information about the temperature within the structure, but they do allow calculation of the specific heat and thermal conductivity. Likewise, molecular dynamics lends itself well to computation of the thermal conductivity at high temperatures, but will also be limited with regard to determination of the associated field, especially if there is a large temperature gradient [72].

In the rest of this section, we shall compare these tools, where possible, and specify their limitations. Calculations concern two semiconductors, silicon and germanium, and also carbon nanotubes.

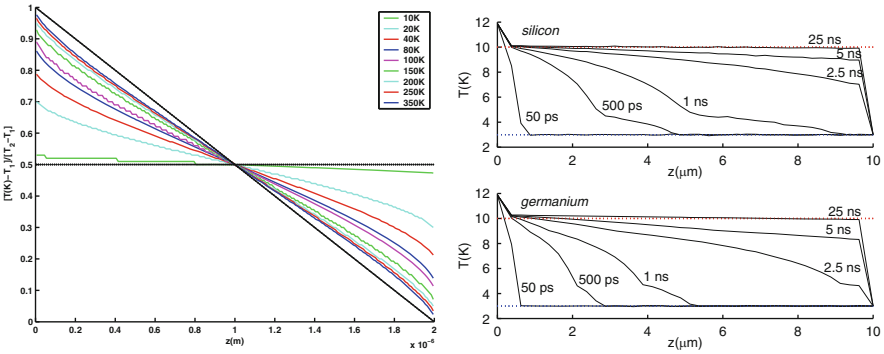

Fig. 2.19 Evolution of the temperature profile in nanofilms. Transition from the diffusive regime to the ballistic regime at different temperatures. *Left*: Silicon film of thickness 2 μm (DOM). *Right*: Silicon and germanium films of thickness 10 μm (MC)

2.4.1 Examples of Confinement in a Nanofilm

Temperature Profiles in Nanofilms

In order to calculate the thermal field in a nanostructure, the Boltzmann transport equation (BTE) must be solved on a spatial lattice. The discrete ordinate and Monte Carlo methods are well suited to this type of calculation. In the case of 1D structures of film type with a temperature imposed on the front and back faces, they model the diffusive regime if the material is thick enough. Furthermore, they model the ballistic regime when the film is very thin or when investigating materials at very low temperatures.

Figure 2.19 (left) shows the variation of T across a film of thickness 2 μm in the steady-state regime. There is a gradual transition toward the ballistic regime as the temperature decreases. This phenomenon results from the increase in the mean free path of the heat carriers as collisions become less common and occur mainly with the boundaries of the structure. Figure 2.19 (right) shows the non-steady state trend toward this ballistic regime reached at low temperatures. Note that there are two modes of heat propagation associated with transverse (TA) and longitudinal (LA) phonons. The latter lead the way owing to their higher group velocity. Moreover, the main part of the energy is transported by transverse phonons, with the two modes interacting very little due to the limited number of collisions. The same type of temporal evolution toward the Fourier regime can be observed without giving rise to these LA and TA waves, since three-phonon collisions changing the frequency and polarisation are much more common.

These two methods DOM and MC for heat transport calculations also turn out to be useful for modelling the non-steady state propagation of a heat pulse through a nanostructure. It is thus possible to envisage modelling the propagation of an acoustic wave in a nanostructure, of the kind produced experimentally in pump–probe experiments.

Thermal Conductivity of Nanofilms

As mentioned in the introduction, calculations of the thermal properties of nanostructured materials, in particular the phonon conductivity, are of great importance in a large number of applications. We shall discuss here some simulation results obtained using the models described earlier.

A prerequisite for modelling the thermal conductivity of a nanostructure is to check that the chosen numerical model can correctly represent this property in the bulk material. In the present case, the three models KT, DOM, and MC satisfy this criterion over a broad temperature range. In the preliminary calculations, the diffusive transport hypothesis is validated and the notion of thermal conductivity is thus relevant. As the characteristic size of the nanostructure decreases, it is often better to switch to the notion of thermal conductance G. For a film of thickness L, the heat transfer in the $\mathbf{e}_z$ direction is associated with conductance G_z given by

$$G_z = \frac{\mathbf{q}_z}{\Delta T} \equiv \frac{k_z}{L} \quad \Longleftrightarrow \quad \text{in the diffusive regime!} \tag{2.60}$$

Note that the notion of thermal conductivity k also involves a directional aspect which is clearly indicated for nanofilms. The conductivity k_z measured across the thickness of the film is called the cross-plane conductivity, while the conductivity k_x or k_y in the plane is specifically referred to as in-plane conductivity.

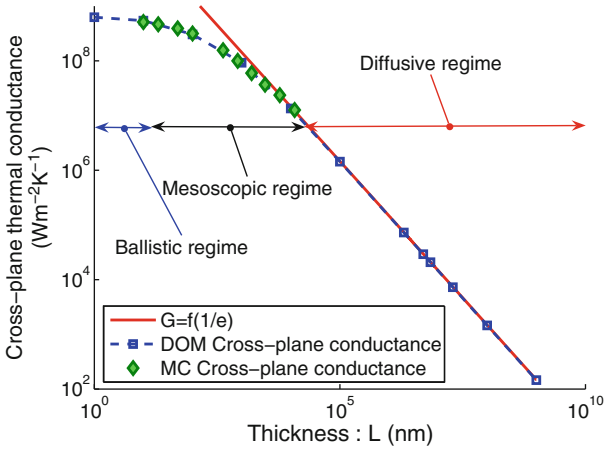

Fig. 2.20 Cross-plane conductance of a silicon film at 300 K

Conduction Regimes in a Nanofilm

To illustrate the above discussion, Fig. 2.20 shows the thickness dependence of the conductance in a silicon film. Three regions of different physical behaviour can be identified. For large values of the thickness, the conductance decreases as $1/L$. Heat transfer is governed by the Fourier law and the thermal conductivity k_z is a relevant quantity. For very small values of the thickness, the conductance saturates and tends toward a constant value. The transport regime here is purely ballistic and the notion of thermal conductivity is no longer meaningful. The intermediate or mesoscopic regime corresponds to the coexistence of both diffusive and ballistic processes within the nanostructure. Once again, the notion of thermal conductivity is somewhat ambiguous.

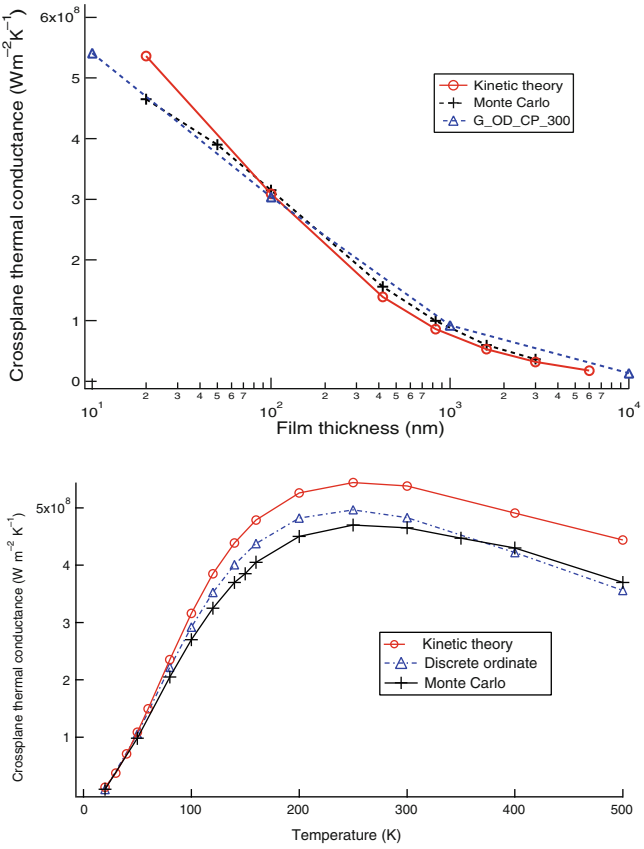

Fig. 2.21 Cross-plane conductance of silicon films. *Top*: Dependence on the thickness L at $T = 400$ K. *Bottom*: Dependence on temperature T for a fixed thickness $L = 20$ nm

Cross-Plane Conductance of a Nanofilm

The conductance of silicon nanofilms was calculated using the KT, DOM, and MC models under similar simulation conditions in order to compare the suitability of the various models. Figure 2.21 (upper) shows the thickness dependence of the conductance over the range 20 nm $< L <$ 6 μm. The values obtained by the three methods are globally rather close. This result, expected for high enough values of the thickness, was rather less so for very thin films. In the case of the KT model, it is the fact that the boundary scattering term τ_B is rigorously taken into account that validates the calculations.

The second comparison we have carried out concerns cross-plane conductance measurements at different temperatures, shown in Fig. 2.21 (lower). Here, too, the agreement between the three numerical models is quite satisfactory. Note the slight overestimate of the values calculated using the KT model, which one must remember is used here in the least favourable conditions (very thin films). The overall trend of the conductance curve for different temperatures T agrees well with theory. At low temperatures ($T <$ 100 K), the conductance varies as $G_z \propto T^3$, while for the highest temperatures ($T >$ 300 K), the dependence is $G_z \propto 1/T$. At intermediate temperatures, it is the increase in the phonon energy and the reduced importance of the resistive umklapp processes that favours heat transfer.

In-Plane Conductivity of a Nanofilm

The in-plane conductivity of a silicon nanofilm was investigated for different thicknesses in the range 20 nm $< L <$ 6 μm (see Fig. 2.22). For this study, only the KT and DOM techniques were used. Solution with the MC method was ruled out because the required computation time would have been too long (2D spatial lattice). In addition, our results are compared with experimental data obtained by M. Ashegi et al. [73], Y.S. Ju and K.E. Goodson [74], and W. Liu and M. Ashegi [75]. The

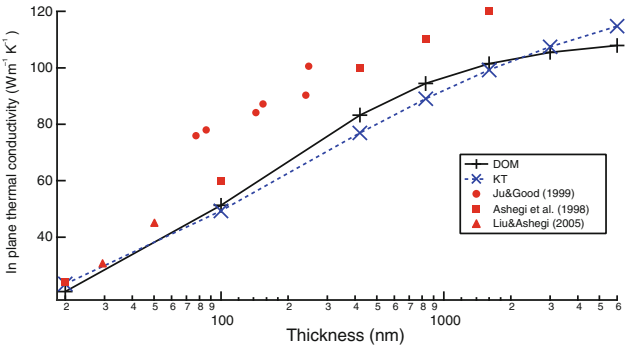

Fig. 2.22 In-plane thermal conductivity of silicon thin films at 300 K. KT model, DOM model, and experimental data

numerical models give equivalent solutions which agree with measurements for small thickness values ($L < 100$ nm). For thicker films, the numerical models underestimate experimental values by some 20%. Since experimental error is of the order of 15%, and since numerical simulation of the experimental conditions is somewhat delicate, these results are globally satisfactory.

2.4.2 Examples of Confinement in a Nanowire and a Nanotube

Silicon Nanowires

The last test case considered concerns silicon nanowires. The thermal conduction properties of these nanostuctures, to be found in particular in new applications to transistors [76], depend sensitively on their dimensions. In contrast to nanofilms, phonon reflection effects at the surface of the nanowire naturally induce resistive mechanisms which oppose heat transport. It is scattering from the whole perimeter of the nanowire that is responsible for this state of affairs. The roughness of the boundaries is taken into account by introducing a certain proportion of diffuse reflection of the phonons at this boundary. This mechanism opposes propagation of heat carriers along the thermal gradient. The calculation results were compared with the experimental data obtained by D. Li et al. [12] for four values of the diameter, viz., $D = 115$ nm, $D = 56$ nm, $D = 37$ nm, and $D = 22$ nm. The simulation results were in good agreement with measured values for the three largest diameters (see Fig. 2.23). As in the previous study, behaviour at low and high temperatures is in agreement with theory. In addition, the thermal conductivity decreases by an order of magnitude in thin wires ($D <37$ nm) at room temperature.

For the wire of diameter $D = 22$ nm, there is disagreement between the numerical and experimental results. The discrepancy may be due to the changed dispersion relations and scattering interaction times caused by confinement of the vibration modes for such small diameters. Moreover, it is clearly a major challenge to carry out thermal conductivity measurements on such small objects.

MD and Carbon Nanotubes

Apart from silicon and germanium nanostructures, a fair number of studies have been made to predict the thermal conductivity of carbon nanotubes (CNT). These are nanostructures with many potential applications in microelectronics [1] and high resolution field emission [2, 3]. The interest of these objects lies in their length to diameter ratio which may attain several orders of magnitude, together with their mechanical, electrical, and thermal properties. Carbon nanotubes are effectively close to being 1D systems. Their thermal conductivity thus depends on the length of the nanotube according to a relation that should be intermediate between logarithmic

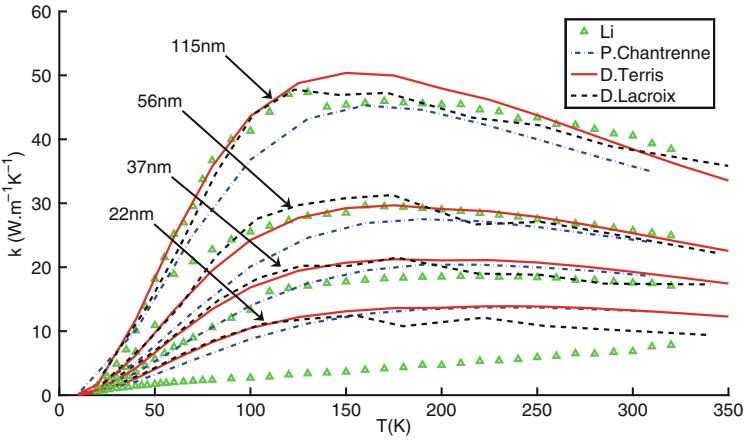

Fig. 2.23 Thermal conductivity of silicon nanowires. KT, DOM, and MC models. Experimental results from Li et al. [12]

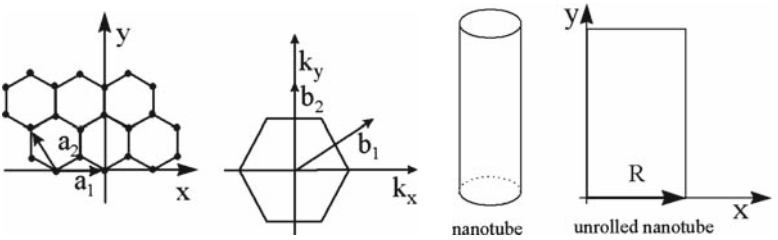

Fig. 2.24 Construction of graphene from the elementary cell. *Left*: In directions x and y, the components of the translation vectors are $\mathbf{a}_1$ $(a_0\sqrt{3}, 0)$ and $\mathbf{a}_2$ $(-a_0\sqrt{3}, a_0 3/2)$, with $a_0 = 1.43$ nm. $\mathbf{b}_1$ and $\mathbf{b}_2$ are the reciprocal vectors of $\mathbf{a}_1$ and $\mathbf{a}_1$. The hexagonal surface is the first Brillouin zone of graphene. *Right*: Definition of the chiral vector of a nanotube

$\ln L$, characterising a 2D system, and the power law $L^{2/5}$, characterising a 1D system [77].

A single-wall carbon nanotube can be viewed as a graphene sheet that has been rolled up on itself. Graphene comprises a hexagonal atomic lattice with unit cell as shown in Fig. 2.24. When a nanotube is unrolled, the vector $\mathbf{R}$ along the edge of the nanotube that has been unrolled in this way, is a linear combination of the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ defining the unit cell: $\mathbf{R} = n\mathbf{a}_1 + m\mathbf{a}_2$, where m and n are the chiral numbers characterising the nanotube. Multiwall nanotubes comprise a superposition of single-wall nanotubes arranged in a layered way like an onion.

Only individual multiwall carbon nanotubes have been characterised experimentally. Values of the thermal conductivity are particularly high [78–80], between 1 000 and 10 000 W/mK depending on the length and diameter of the nanotube, except for nanotubes including a high concentration of defects [81,82]. Unfortunately, these nanotubes are too big to be accessible to molecular dynamics predictions of their thermal conductivity, requiring excessive computer capacity. Hone et al. [83] have estimated the thermal conductivity of single-wall carbon nanotubes with values

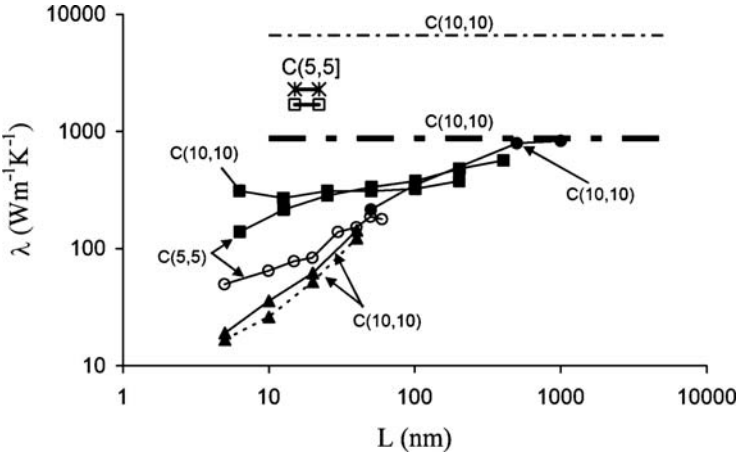

Fig. 2.25 Comparison of the thermal conductivity of C(5,5) and C(10,10) carbon nanotubes at 300 K, obtained from several MD simulations. *White circles*: Our results using NEMD. *Black squares*: Results from [47] using NEMD. *Black triangles* are results from [46] using EMD without correction (*continuous line*) or with temperature correction (*dashed line*). *Black circles*: Results from [50] using NEMD. *Thick dot-dashed line*: Results obtained from EMD [49] (length unspecified). *Normal dot-dashed line*: Results from HEMD [48] (length unspecified). *Stars* and *white squares*: Results from [51] (length some tens of nm, not precisely specified)

between 1 700 and 5 800 W/mK, by measuring the thermal conductivity of a rope made from single-wall nanotubes. In this case, the inaccuracy of the result excludes reliable comparison with predictions.

Predictions of the thermal conductivity of carbon nanotubes appeal to an application of molecular dynamics. As the interaction potential between the carbon atoms is rather complex, this leads to prohibitive computation times and only small single-wall nanotubes of type C(5,5) and C(10,10) have been considered [46–51, 84]. The three simulation methods, EMD, NEMD, and HNEMD, have been used. Despite a significant scatter in the results (see Fig. 2.25), the orders of magnitude are consistent, excepting the results of Yao et al. [84], not shown in the figure, which are an order of magnitude greater than all the other results. The thermal conductivity does indeed depend on the length through a power law.

By considering the vibrational properties of graphene, the model parameters from the kinetic theory of gases have been determined in order to predict the thermal conductivity of carbon nanotubes. The results are compared with those from non-equilibrium molecular dynamics for C(5,5) nanotubes as a function of temperature for a nanotube 10 nm long (see Fig. 2.26) and as a function of the length for a temperature of 300 K (see Fig. 2.27). The molecular dynamics results are not corrected, as is the case for those proposed by Lukes et al. A comparison between the two models is thus only justified if the model from the kinetic theory of gases treats a uniform distribution of energy over all vibration modes. When this is the case, the results are in good agreement. In addition, both models imply the expected dependence of the thermal conductivity on the length of the nanotube.

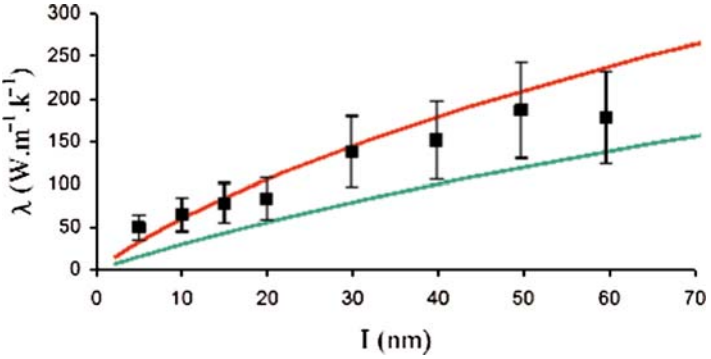

Fig. 2.26 Length dependence of the thermal conductivity of a C(5,5) carbon at 300 K. *Symbols*: Molecular dynamics results with the non-equilibrium method. *Red*: Results of the model based on the kinetic gas theory with constant energy for each vibrational mode. *Green*: Results of the model based on the kinetic gas theory and taking into account quantisation of the phonon energy

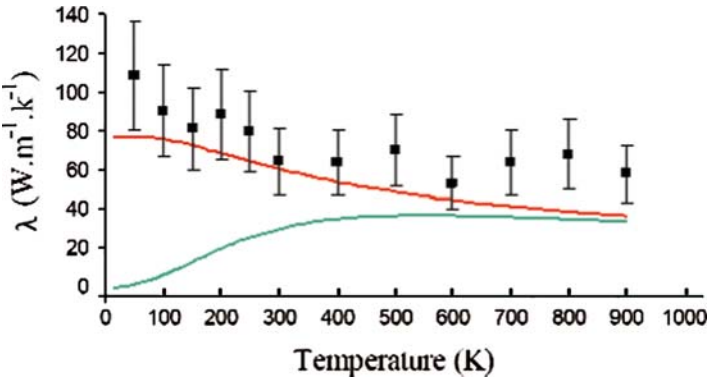

Fig. 2.27 Temperature dependence of the thermal conductivity of a C(5,5) carbon nanotube of length 10 nm. *Symbols*: Molecular dynamics results with the non-equilibrium method. *Red*: Results of the model based on the kinetic gas theory with constant energy for each vibrational mode. *Green*: Results of the model based on the kinetic gas theory and taking into account quantisation of the phonon energy

2.5 Conclusion

The aim in this chapter has been a practical discussion of models capable of describing heat exchange in nanostructures. This information together with the extensive bibliography to follow should constitute as exhaustive a foundation as possible for applying the tools of computer simulation to heat transfer within nanostuctures.

Section 2.2.1 dealt with the physics required to understand the different models, describing the various phenomena associated with heat transport, with emphasis on the notion of phonon. We then discussed the semi-analytic models widely used to calculate the heat conduction properties of nanostructures (Sect. 2.2.2).

Section 2.3 described the main numerical techniques for representing heat transfer, i.e., deterministic, statistical, and mechanical methods. The aim of this section

was not to place one above the other, but rather to bring out their respective advantages and disadvantages. Moreover, as shown in the last section (Sect. 2.4) which describes some of the results produced by the various approaches, when the conditions of use are comparable, the different methods often produce very similar results.

At the present time, a lot of work is being done on the faithful representation of the heat transfer properties of nanostructures. In many cases, the goals are clearly established, e.g., the fabrication of thermoelectric materials, the optimisation of microelectronic components, or the improvement of photovoltaic components. In others, studies are carried out on the physics of transport to achieve a better understanding of the interaction mechanisms associated with phonons. This is particularly true of studies that aim to provide a better definition of the coupling between acoustic and optical phonons, or those that seek to model the vibrational properties associated with specific nanostructures like quantum dots, quantum wells, nanowires, and many others.

The field of investigation is still considerable. For an area of research which takes its roots in the work of Debye and Peierls at the beginning of the last century, much still remains to be done!

Appendix: Measuring Thermal Properties

This section focuses on experimental results concerning the thermal conductivity of nanostructures. Nanostructured materials such as nanoporous materials, nanocomposites, nanofluids, superlattices, or nanosequenced materials will be considered later in the book. The aim of this review is to give an overview of the measurements that have been made and which have been used to validate the theoretical models.

Several kinds of classification may be considered, depending on:

- the nature of the material, e.g., electric or dielectric,
- the geometry of the nanostructure, e.g., nanoparticle, nanotube, nanowire, nanofilm, etc.
- the composition of the material, e.g., Si, SiO_2, SiGe, Ge, gold, platinum, carbon, etc.

Due to the wide use of nanostructures in microelectronics, the thermal conductivities of many nanostructures made of dielectrics have been measured. Among them, nanotubes, nanowires, and nanofilms are the most numerous.

The thermal conductivities of silicon nanofilms with thicknesses 420 nm, 830 nm, and 1.6 μm have been measured for temperatures varying between 10 and 400 K [73]. Si films 20 and 100 nm thick were characterised in 2004 for a temperature range of 20 to 300 K [85].

In 2003, Li and al. [12] measured the thermal conductivity of Si nanowires with typical diameters of 22, 37, 56, and 115 nm for temperatures ranging between 20 and 320 K. More recently, Zang and al. [86] used thermoreflectance to measure the thermal conductivity of a silicon nanowire of diameter 115 nm and length

3.9 μm. The value is only 15% higher than the one obtained previously. At very low temperatures, studying Si nanowires with length 5 to 15 μm and cross-section $130 \times 200\ \text{nm}^2$, Bourgeois et al. [87] showed that the expected T^3 variation of the thermal conductivity fails below 1.4 K.

The thermal conductivity of a 190 nm thick SiO_2 film has been measured by Mavrokefalos et al. [88] and compared with the measurements made by Cahill et al. [89] years earlier on a 900 nm thick SiO_2 film. From 300 to 450 K, the former values are 30% lower than the latter. This result is unexpected since SiO_2 is amorphous and the phonon mean free path is much smaller than the film thicknesses. The discrepancy might be explained by differences in the method of fabrication.

The thermal conductivity of SWNTs of length 1 μm and diameter 1.4 nm has been estimated from the measurement of a crystalline rope of SWNT [83]. The thermal conductivities of several MWNTs were measured at different temperatures [78]. The thermal conductivity of an MWNT 40 μm long and 10 nm in diameter was estimated by comparing a heat transport model and experimental measurements of the temperature profile of this MWNT during field emission. The low value of the thermal conductivity obtained with this method may be explained by the high defect concentration of the MWCNT used in these experiments [81, 82]. More recently, the thermal conductivities of several multiwall carbon nanotubes were measured by Fujii et al. [79] using another microelectronic setup. They confirm the order of magnitude and the large temperature variations of the thermal conductivity of MWCNTs.

Metallic films are often parts of microelectronic systems used to characterize nanostructures. A knowledge of their thermal and electrical properties is then also important. It has been found that the thermal conductivity of a thin film might be lower than the bulk value. As an example, Zhang et al. [90] have shown that the thermal conductivity of gold nanofilms increases with film thickness, when the latter varies from 21 to 37 nm, and is about 40–50% smaller than the bulk value. The variation in thermal conductivity is correlated with the grain size, which also increases with the film thickness. Xing and al. [91] worked on Pt nanofilms. For a Pt thin film with width 260 nm, thickness 28 nm, and length 5.3 μm, the in-plane thermal conductivity is of the order of 25–30% of the corresponding bulk value from 77 to 330 K. The low values of the thermal conductivity for thin films may be attributable to structural defects, surface scattering, grain boundary scattering, film sizes, fabrication methods, and so on.

References

1. P.C. Collins, M.S. Arnold, and P. Avouris: Engineering carbon nanotubes and nanotube circuits using electrical breakdown, Science **292**, 706 (2001)
2. W. de Heer, A. Chatelain, and D. Ugarte: A carbon nanotube field-emission electron source, Science **270**, 1179 (1995)
3. J.M. Bonard, H. Kind, T. Stckli, and L. O Nilsson: Field emission from carbon nanotubes: The first five years, Solid State Electron. **45** (6), 893–914 (2001)

4. C. Kittel: *Introduction to Solid State Physics*, 8th edn., John Wiley, Philiadelphia, 2004
5. N.W. Ashcroft and N.D. Mermin: *Solid State Physics*, International Edition, Saunders College, Philiadelphia, 1976
6. R. Turbino, L. Piseri, and G. Zerbi: Lattice dynamics and spectroscopic properties by a valence force potential of diamondlike crystals: C, Si, Ge, and Sn, J. Chem. Phys. **56**, 1022–1039 (1972)
7. E. Pop, R.W. Dutton, and K.E. Goodson: Analytic band Monte Carlo model for electron transport in Si including acoustic and optical phonon dispersion, J. Appl. Phys. **96**, 4998–5005 (2004)
8. C. Dames and G. Chen: J. Appl. Phys. **89**, 682–693 (2003)
9. Y.-J. Han and P.G. Klemens: Phys. Rev. B **48**, 6033–6042 (1993)
10. A. Balandin and K.L. Wang: Phys. Rev. B **58**, 1544–1549 (1998)
11. N. Mingo: Phys. Rev. B **68**, 113308 (2003)
12. D. Li, Y. Wu, P. Kim, P. Yang, and A. Majumdar: Appl. Phys. Lett. **83**, 2934–2936 (2003)
13. P.G. Klemens: The scattering of low-frequency lattice waves by static imperfections, Proc. Phys. Soc. A **68** (12), 1113–1128 (1955)
14. P.G. Klemens: Proc. Roy. Soc. A **208**, 108–133 (1951)
15. J. Callaway: Model for lattice thermal conductivity at low temperatures, Phys. Rev. **113**, 1046–1051 (1959)
16. M.G. Holland: Analysis of lattice thermal conductivity, Phys. Rev. **132** (6), 2461–2471 (1963)
17. P. Chantrenne and J.L. Barrat: Finite size effects in determination of thermal conductivities: Comparing molecular dynamics results with simple models, J. Heat. Transf. **126**, 577–585 (2004)
18. P. Chantrenne, J.L. Barrat, X. Blase, and J.D. Gale: J. Appl. Phys. **97**, 104318 (2005)
19. E. Akkermans and G. Montambaux: *Physique mésoscopique des électrons et des photons*, EDP Sciences, Paris, 2004
20. G. Chen: Phys. Rev. B **23**, 14958–14973 (1998)
21. K. Schwab, E.A. Henriksen, J.M. Worlock, and M.L. Roukes: Measurement of the quantum of thermal conductance, Nature (London) **404**, 974–976 (2000)
22. S.G. Walkauskas, D.A. Broido, K. Kempa, and T.L. Reinecke: Lattice thermal conductivity of wires, J. Appl. Phys. **85** (5), 2579–2582 (1999)
23. J. Zou and A. Balandin: Phonon heat conduction in a semiconductor nanowire, J. Appl. Phys. **89** (5), 2932–2938 (2001)
24. X. Lu and J. Chu: J. Appl. Phys. **100**, 14305 (2006)
25. X. Lu, Z. Shen, and J.H. Chu: J. Appl. Phys. **91**, 1542 (2002)
26. M.F. Modest: *Radiative Heat Transfer*, 2nd edn., Academic Press, San Diego, 2003
27. S. Chandrasekhar: *Radiative Transfer*, Dover, New York, 1960
28. S. Volz (Ed.): *Microscale and Nanoscale Heat Transfer*, Topics in Applied Physics Vol. 107, Springer, Berlin Heidelberg, 2007
29. R.B. Peterson: J. Heat. Transf. **116**, 815–822 (1994)
30. S. Mazumder and A. Majumdar: Monte Carlo simulations of phonon transport in solid thin films including dispersion and polarization, J. Heat. Transf. **123**, 749–759 (2001)
31. D. Lacroix, K. Joulain, and D. Lemonnier: Monte Carlo transient phonon transport in silicon and germanium at nanoscales, Phys. Rev. B **72**, 064305 (2005)
32. D. Lacroix, K. Joulain, D. Terris, and D. Lemonnier: Monte Carlo transient phonon transport in silicon and germanium at nanoscales, Appl. Phys. Lett. **89**, 103104 (2006)
33. D.J. Oh and R.A. Johnson: J. Mater. Res. **3** (3), 471–478 (1988)
34. F. Ercolessi, M. Parrinello, and E. Tosatti: Phil. Mag. **58** (1), 213–226 (1988)
35. F.H. Stillinger and T.A. Weber: Phys. Rev. B **31** (8), 5262–5271 (1985)
36. R. Biswas and D.R. Hamann: Phys. Rev. Lett. **55** (19), 2001–2004 (1985)
37. D.W. Brenner: Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films, Phys. Rev. B **42** (15), 9458–9471 (1990)
38. J. Tersoff: Phys. Rev. B **37** (12), 6991–7000 (1988)

39. F. Ercolessi and J.B. Adams: Europhys. Lett. **26** (8), 583–588 (1994)
40. M.P. Allen and D.J. Tildesley: *Computer Simulation of Liquids*, Oxford University Press, New York, 1997
41. D. Frenkel and S. Berend: *Understanding Molecular Simulation*, Academic Press, San Diego, 1996
42. D.C. Rapaport: *The Art of Molecular Dynamics Simulation*, Cambridge University Press, Cambridge, 1995
43. R.J. Sadus: *Molecular Simulation of Fluids*, Elsevier Science, Amsterdam, 1999
44. P. Chantrenne and S. Volz: Etude par dynamique moléculaire: Introduction. *Techniques de l'Ingénieur*, 126:BE 8290, 2002
45. S. Volz and P. Chantrenne: Etude par dynamique moléculaire; conduction thermique aux nanoéchelles. *Techniques de l'Ingénieur*, page BE 8291, 2002
46. J.R. Lukes and Z. Hongliang: Thermal conductivity of individual carbon nanotubes, J. Heat. Transf. **129**, 705 (2007)
47. S. Maruyama: Molecular dynamics simulation of generation process of SWNTs, Physica B **323** (1), 187–189 (2002)
48. S. Berbert, Y.K. Kwon, and D. Tomanek: Unusually high thermal conductivity of carbon nanotubes, Phys. Rev. Lett. **84** (20), 4613–4616 (2000)
49. J. Che, T. Cagin, and W.A. Goddard: Thermal conductivity of carbon nanotubes, Nanotechnology **11**, 65–69 (2000)
50. J.F. Moreland: The disparate thermal conductivity of carbon nanotubes and diamond nanowires studied by atomistic simulations, Nanoscale and Microscale Thermophysical Engineering **8** (1), 61–64 (2004)
51. M.A. Osman and D. Srivastava: Temperature dependence of the thermal conductivity of single wall carbon nanotubes, Nanotechnology **12**, 21–24 (2001)
52. J.R. Lukes, D.Y. Li, X.G. Liang, and C.L. Tien: J. Heat. Transf. **122**, 536 (2000)
53. C.J. Gomes, M. Madrid, J.V. Goicochea, and C.H. Amon: In-plane and out-of-plane thermal conductivity of Si thin films predicted by molecular dynamics, J. Heat. Transf. **128**, 1114–1121 (2006)
54. S.P. Yuan and P.X. Jiang: Thermal conductivity of small nickel particles, International Journal of Thermophysics **27** (2), 581–595 (2006)
55. S. Chapman and T.G. Cowling: *The Mathematical Theory of Non-Uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*, Cambridge University Press, Cambridge, 1970
56. B. Diu, C. Guthmann, D. Lederer, and B. Roulet: *Eléments de physique statistique*, Hermann, Paris, 1989
57. R. Zwanzig: Ann. Rev. Phys. Chem. **16**, 67 (1964)
58. A.J.C. Ladd, B. Moran, and W.G. Hoover: Phys. Rev. B **34**, 5058 (1996)
59. R. Kubo, M. Toda, and N. Hashitsume: *Statistical Physics 2, Nonequilibrium Statistical Mechanics*, 2nd edn., Springer, Berlin, 1991
60. F. Müller-Plathe: J. Chem. Phys. **106**, (14), 6082 (1997)
61. P. Jund and R. Jullien: Phys. Rev. B **59** (21), 13707–13711 (1999)
62. T. Ikeshoji and B. Hafskjold: Mol. Phys. **81**, 251–261 (1994)
63. B.C. Daly and H.J. Maris: Physica B **316–317**, 247–249 (2002)
64. J.L. Barrat and F. Chiaruttini: Kapitza resistance at the liquid–solid interface, Mol. Phys. **101**, 15605 (2003)
65. J.C. Schön C. Olischlger: Phys. Rev. B **59** (6), 4125–4133 (1999)
66. P. Chantrenne and J.L. Barrat: Eurotherm seminar 75 on microscale heat transfer. In: *Physics of Defects*, Reims, France, 2003
67. T. ZhenAn and H. ZhengXing: Evaluation of momentum conservation influence in non-equilibrium molecular dynamics methods to compute thermal conductivity, Physica B **373** (2), 291–296 (2006)
68. D.J. Evans: Phys. Lett. A **91** (9), 457 (1982)
69. D.J. Evans: Phys. Rev. A **34** (2), 1449 (1986)

70. D.J. Evans: J. Chem. Phys. **78**, 3297 (1983)
71. D.J. Evans, W.G. Hoover, B.H. Failor, B. Moran, and A.J.C. Ladd: Phys. Rev. **28**, 1016 (1983)
72. S. Volz and G. Chen: Molecular-dynamics simulation of thermal conductivity of silicon crystals, Phys. Rev. B **61**, 2651–2656 (2000)
73. M. Asheghi, M.N. Toulzebaev, K.E. Goodson, Y.K. Leung, and S.S. Wong: J. Heat. Transf. **120**, 30 (1998)
74. Y.S. Ju and K.E. Goodson: Appl. Phys. Lett. **74** (20), 3005–3007 (1999)
75. W. Liu and M. Asheghi: Thermal conductivity in ultra thin pure and doped single crystal silicon layers at high temperatures. In: *Proc. ASME Summer Heat Transfer Conference*, pp. HT2005–72540, San Francisco, USA, 2005
76. J. Chen and R. Konenkamp: Vertical nanowire transistor in flexible polymer foil, Appl. Phys. Lett. **82**, 4782–4784 (2003)
77. S. Lepri, R. Livi, and A. Politi: Thermal conduction in classical low-dimensional lattices, Phys. Rep. **377**, 1–80 (2003)
78. P. Kim, L. Shi, A. Majumdar, and P.L. Mc Euen: Thermal transport measurements of individual multiwalled carbon nanotubes, Phys. Rev. Lett. **87**, 215502 (2001)
79. M. Fujii, X. Zhang, H. Xie, H. Ago, K. Takahashi, T. Ikuta, H. Abe, and Tetsuo Shimizu: Measuring the thermal conductivity of a single carbon nanotube, Phys. Rev. Lett. **95**, 065502 (2005)
80. P. Kim, L. Shi, A. Majumdar, and P. L. McEuen: Mesoscopic thermal transport and energy dissipation in carbon nanotubes, Physica B **323**, 67–70 (2002)
81. P. Vincent, S.T. Purcell, C. Journet, and V.T. Binh: Modelization of resistive heating of carbon nanotubes during field emission, Phys. Rev. B **66**, 075406 (2006)
82. S.T. Purcell, P. Vincent, C. Journet, and V.T. Binh: Hot nanotubes: Stable heating of individual multiwall carbon nanotubes to 2000 K induced by the field-emission current, Phys. Rev. Lett. **88** (10), 105502 (2000)
83. J. Hone, M. Whitney, C. Piskoti, and A. Zettl: Thermal conductivity of single-walled carbon nanotubes, Phys. Rev. B **59** (4), R2514–R2516 (1999)
84. Z. Yao, J. Wang, B. Li, and G. Liu: Thermal conduction of carbon nanotubes using molecular dynamics, Phys. Rev. B **71**, 085417 (2005)
85. W. Liu and M. Asheghi: Appl. Phys. Lett. **84** (19), 3819–3821 (2004)
86. Y. Zhang, J. Christofferson, A. Shakouri, D. Li, A. Majumdar, Y. Wu, R. Fan, and P. Yang: IEEE Transactions on Nanotechnology **5** (1), 67–74 (2006)
87. O. Bourgeois, T. Fournier, and J. Chaussy: J. Appl. Phys. **101**, 16104 (2007)
88. A. Mavrokefalos, M.T. Pettes, Z. Feng, and S. Li: Rev. Sci. Instrum. **78**, 34901/1–6 (2007)
89. D.G. Cahill, M. Katiyar, and J.R. Abelson: Phys. Rev. B **50**, 6077–6081 (1994)
90. Q.G. Zhang, B.Y. Cao, X. Zhang, M. Fujii, and K. Takahashi: Phys. Rev. B **74** (13), 134109 (2006)
91. Z. Xing, X. Huaqing, M. Fujii, H. Ago, K. Takahashi, T. Ikuta, H. Abe, and T. Shimizu: Appl. Phys. Lett. **86** (17), 1719126 (2006)

Chapter 3
Green's Function Methods for Phonon Transport Through Nano-Contacts

Natalio Mingo

3.1 Introduction to Green's Functions for Lattice Thermal Transport

Formulations of the phonon transport problem depend on the length scale of interest. At the bottom end of the scale is the atomistic description. There are fundamental reasons for going to the atomic description level. One is that atomistic calculations can be used to extract parameters that are needed for coarser-grained descriptions. For example, descriptions like the equation of phonon radiative transfer [1,2], or at a coarser level, the heat diffusion equation, need to be provided with magnitudes such as interface thermal resistance and local thermal conductivity tensors. To obtain such local properties, atomistic approaches have been developed. Some examples are the Kubo formula for molecular dynamics, which yields the bulk thermal conductivity [3], the Allen–Feldman approach for the thermal conductivity of amorphous solids [4], and the method of lattice dynamics for interface thermal resistance [5–7].

In many cases, experimental validation of these local properties can only be carried out indirectly, by measurements on macroscopic samples. Nevertheless, experimental techniques have recently been developed that allow local thermal transport measurements on samples with characteristic dimensions in the sub-micrometer range [8–10]. Such 'local probe' experiments are generally based on measuring the heat flow across a nanoscale object linking two heat reservoirs which are kept at two different temperatures (see Fig. 3.1).

The kind of experiment depicted in Fig. 3.1 is best suited for a Green's function formulation. This is so for several reasons. First, Green's functions allow one to treat semi-infinite, non-periodic systems. Thus, no spurious periodicities need to be imposed. A great advantage of Green's function methods is that one can effectively 'project' infinitely large parts of the system onto the part one is interested in. This is extremely useful when modeling the heat reservoirs which supply the non-equilibrium phonons flowing through the system. Another reason is that, when the size of a system is comparable or shorter than the bulk mean free path, phonons

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_3, © Springer-Verlag Berlin Heidelberg 2009

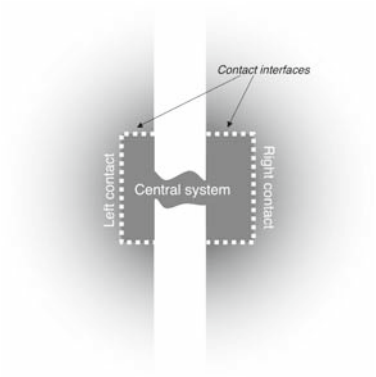

Fig. 3.1 Nano-bridge linking two systems at different temperatures. Phonon transport through this type of structure is suitable for study by Green's function techniques

travel nearly ballistically, and anharmonic effects are less important. This simplifies the problem enormously, since it reduces to one of non-interacting phonons. In a purely harmonic case, Green's functions have advantages over a molecular dynamics simulation.

Molecular dynamics [11] represents an alternative to Green's function techniques. The two most fundamental differences between the two, physically speaking, are:

- GF is quantum mechanical whereas MD is classical and thus restricted to high temperatures.
- MD includes anharmonic interactions to all orders, whereas GF is presently limited to length scales at which anharmonicity is of secondary importance. (A way to include anharmonicity perturbatively is presented here in Sect. 3.3.)

Therefore, despite their radically different formulations, the two techniques can complement one another, each of them working best in the cases where the other one might fail. It is thus important to carry out studies to compare the outcomes from both types of calculation. However, not much appears to have been done so far in this direction.

This chapter is divided into two parts. The first one deals with the theory of the harmonic problem. This part has been written keeping students in mind. One important goal has been to derive the basic formalism in a simple way. Thus, advanced concepts such as those from many-body theory have been purposely avoided. Some simple worked problems are provided for the reader to get acquainted with the basic theoretical concepts, like the projection of a semi-infinite system via self-energies, or the calculation of transmission functions. A miscellaneous collection of applications of the transmission function concept to real systems has also been included.

The second part deals with the formulation of the anharmonic phonon transport problem. The material in this part is new to a large extent. The second part is much more specialized than the first, and it has been written for more advanced readers.

Some familiarity with many-body concepts is needed. The chapter ends with a non-exhaustive summary of previous work, together with a discussion of some important problems and directions for future work.

3.2 The Harmonic Problem

3.2.1 Dynamics of Non-Periodic Systems

This section is a brief introduction to some concepts of harmonic lattice dynamics and single-particle Green's functions. A good and more extensive introduction to these concepts can be found in [12].

The motion of a generic harmonic system is completely determined by its force constant matrix K. This matrix corresponds to the second derivatives of the total energy of the system, E, with respect to any pair of degrees of freedom, u_i and u_j:

$$\tilde{K}_{ij} = \frac{\partial^2 E}{\partial u_i \partial u_j} \,. \tag{3.1}$$

For the Newtonian momentum conservation condition to be fulfilled, the matrix $\tilde{\mathsf{K}}$ satisfies various sum rules, essential when checking the correctness of the force constant matrix [13].

The vibrational normal modes of the system, $\overline{\psi}^{(n)} \mathrm{e}^{-\mathrm{i}\omega_n t}$, satisfy the equation

$$M_i \omega_n^2 \psi_i^{(n)} - \sum_j \tilde{K}_{ij} \psi_j^{(n)} = 0 \,. \tag{3.2}$$

Here, M_i is the mass of the atom to which degree of freedom i belongs. From now on, a bar on top of a symbol will denote that the symbol is a rank 1 tensor, while symbols like K, I, etc., will denote rank 2 tensors (matrices). With this notation, the above equation can be transformed into the matrix dynamical equation:

$$\left(\omega_n^2 \mathsf{I} - \mathsf{K}\right) \overline{\psi}^{(n)} = \bar{0} \,, \tag{3.3}$$

where we have defined the harmonic matrix: $K_{ij} = (M_i M_j)^{-1/2} \partial^2 E / \partial u_i \partial u_j$.

The resolvent Green's function matrix for the system is defined as

$$\mathsf{G}(\omega^2) = \left[(\omega^2 + \mathrm{i}\delta)\mathsf{I} - \mathsf{K} \right]^{-1} , \tag{3.4}$$

with $\delta \to 0^+$. A very important property that we will use in later sections relates the resolvent and the local spectral density of states, ρ. For this, we express G in the basis of eigenstates of K:

$$G_{ij}(\omega) = \sum_n \frac{\psi_i^{(n)*}\psi_j^{(n)}}{\omega^2 + \mathrm{i}\delta - \omega_n^2} \,, \tag{3.5}$$

where the $\overline{\psi}^{(n)}$ form a complete orthonormal set. Then the difference between the resolvent and its transpose conjugate is

$$G_{ij} - G_{ji}^* = \sum_n \psi_i^{(n)*}\psi_j^{(n)} 2\pi\delta(\omega^2 - \omega_n^2) \equiv 2\pi\rho_{ij}(\omega) \,, \tag{3.6}$$

which is just 2π times the local spectral density of states.

The resolvent is an extremely useful tool to deal exactly with the effect of perturbations to the system. If in addition to the eigenstates one knows the resolvent for the system, then one can compute how the eigenfunctions are modified when the force constant matrix is modified on a certain region, for example, by compressing part of the system, establishing new bonds, creating defects, or any other kind of change to the atomic configuration. Let K_0 and $\{\overline{\phi}^{(n)}\}$ be the harmonic matrix and eigenfunctions of the original system, and $\mathsf{K}_0 + \mathsf{V}$ the harmonic matrix corresponding to the modified system. The eigenfunctions $\{\overline{\psi}^{(n)}\}$ of the modified system are related to the original ones by the Lippmann–Schwinger equation [12]:

$$\overline{\psi}^{(n)} = \overline{\phi}^{(n)} + \mathsf{g}(\omega_n^2)\mathsf{V}\overline{\psi}^{(n)} \,, \tag{3.7}$$

where g is the resolvent of the original system. Another useful equation that can be derived is [12]:

$$\overline{\psi}^{(n)} = \overline{\phi}^{(n)} + \mathsf{G}(\omega_n^2)\mathsf{V}\overline{\phi}^{(n)} \,, \tag{3.8}$$

where G is now the resolvent of the modified system.

In the actual computation of the resolvent, however, the definition (3.4) is seldom used by itself. Doing so would involve inverting an extremely large, or ideally, an infinitely large matrix. The advantage of the resolvent is that we can compute its elements on a selected region of the system, and use those elements to obtain properties of the system. The way to compute local elements of the resolvent is to use projection techniques. A central relation for most of these techniques is the Dyson equation, which, similarly to the Lippmann–Schwinger equation for the eigenstates, relates the resolvent G of the modified system to the resolvent g of the original system [12]:

$$G_{ij} = g_{ij} + \sum_{lm} g_{il} V_{lm} G_{mj} \,. \tag{3.9}$$

Since V_{lm} is non-zero only for a finite set of degrees of freedom on a particular region, it suffices to know g_{ij} on that region in order to obtain G.

Various projection techniques have been developed in recent decades, each of them being best suited for a specific kind of physical system. All these techniques were originally developed with the electronic structure problem in mind. Nevertheless, their application to the phonon problem is straightforward. The goal, in all cases, is to obtain the resolvent in a finite region of an infinitely extended, atomically

defined system. In the electron case, the system is described by a tight-binding Hamiltonian. To apply any of these techniques to the lattice vibration case, it suffices to replace the electron energy E by the squared frequency ω^2, and the Hamiltonian matrix H by the harmonic matrix K.

We will comment briefly on two important projection techniques, well-suited to two different kinds of system. Details on how to implement them can be found in the literature. The first is the decimation technique [14, 15]. This technique is designed to evaluate the resolvent at the surface of a semi-infinite crystal. The method is based on a renormalization procedure. It consists in computing the resolvent at the surface layer of atoms of a slab with finite thickness. The slab thickness is doubled at each iteration, until the computed resolvent does not change from one iteration to the next (implying that the system is thick enough to behave like a semi-infinite solid). Because of the doubling procedure, convergence is very fast, usually being achieved in under a dozen iterations. There are different implementations in the literature. The original reference [14] was based on a computation of the self-energy, from which the resolvent is obtained later. On the other hand, [15] directly computes the resolvent at each iteration.

The second projection technique is the recursion method [16]. This technique is best suited to obtaining the resolvent of an atomic region A surrounded by non-crystalline material. Using orthogonalization procedures, the method transforms the system into an equivalent linear chain of inequivalent systems (or sites), where the edge site corresponds to region A. The chain is constructed one step at a time, and it is truncated at the point when adding one more site does not change the resolvent of its edge site. The solution is computationally more efficient than other methods based on adding concentric shells.

Problem 1. Obtain the resolvent on the last atom of a one-dimensional atomic chain with first neighbor interactions.

Solution. Let us take the atomic mass to be 1. The atoms are labeled as shown in Fig. 3.2, the last atom being labeled as number 1. The harmonic matrix is

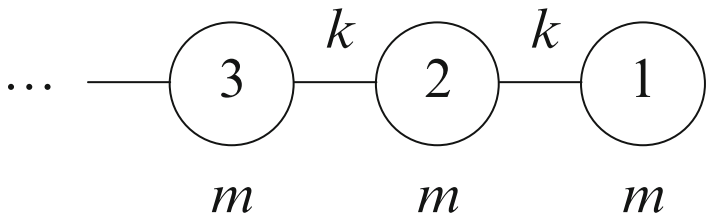

Fig. 3.2 Configuration in Problem 1. A semi-infinite one-dimensional chain with nearest neighbor spring constants k, and atomic masses m

$$K_{ij} = \begin{cases} (\delta_{i,j+1} + \delta_{i,j-1})k - 2k\delta_{ij} \,, & i,j \neq 1 \,, \\ -k \,, & i = 1 = j \,. \end{cases} \tag{3.10}$$

Let us write K in the form $\mathsf{K} = \mathsf{K}^{(0)} + \mathsf{V}$, with $V_{1,2} = V_{2,1} = k$, and any other elements of V equal to 0. In this case, the Dyson equation (3.9) is, for G_{11},

$$G_{11} = g_{11} + g_{11}kG_{21} \,, \tag{3.11}$$

and for G_{21},

$$G_{21} = g_{22}kG_{11} \,. \tag{3.12}$$

In the latter, $g_{21} = 0$, because $K^{(0)}_{12} = 0$, i.e., $\mathsf{K}^{(0)}$ does not connect atom 1 to the rest of the system. Also because of this, $g_{11}(z) = 1/(z+k)$, where $z \equiv \omega^2 + \mathrm{i}0^+$. The above two equations then imply

$$G_{11} = \frac{g_{11}}{1 - k^2 g_{11} g_{22}} = \frac{1}{z + k - k^2 g_{22}} \,. \tag{3.13}$$

It only remains to calculate g_{22}. For this, we can use the Dyson equation again, by noticing that if we decouple atoms 2 and 3, by expressing $K^{(0)} = \tilde{K} + \tilde{V}$ with $\tilde{V}_{ij} = k(\delta_{i,2}\delta_{j,3} + \delta_{i,3}\delta_{j,2})$, then $\tilde{K}_{ij} = K^{(0)}_{i-1,j-1}$ for $i,j > 2$. If $\tilde{\mathsf{g}}$ denotes the resolvent of the system corresponding to matrix $\tilde{\mathsf{K}}$, we have $\tilde{g}_{33} = g_{22}$. The Dyson equation then yields

$$g_{22} = \tilde{g}_{22} + \tilde{g}_{22}kg_{32} = \tilde{g}_{22} + \tilde{g}_{22}k^2 g_{22}^2 \,. \tag{3.14}$$

Since $\tilde{g}_{22} = 1/(z+2k)$, (3.14) yields the following quadratic equation for g_{22}:

$$k^2 g_{22}^2 - (z+2k)g_{22} + 1 = 0 \,, \tag{3.15}$$

with solution

$$g_{22} = \frac{z + 2k + \sqrt{z^2 + 4kz}}{2k^2} \,. \tag{3.16}$$

Thus finally, from (3.13),

$$G_{11} = \frac{2}{\omega^2 - \sqrt{\omega^4 + 4k\omega^2}} \,. \tag{3.17}$$

3.2.2 The Heat Current

The typical configuration of non-equilibrium problems treated with Green function techniques is shown in Fig. 3.1. Basically, the system consists of two semi-infinite reservoirs at two different temperatures, joined by a central system. This central system can either be singly connected at only a finite region (for example a 'point contact'), or it can be continuously connected everywhere along the interface. In the latter case, the problem is formulated in a hybrid representation, where the two

directions parallel to the interface are treated in the reciprocal space. This approach has been implemented in [17]. The following discussion will focus on the singly connected geometry. Details of the continuously connected case can be found in [17].

Let us consider an arbitrary vibrational wave (in general, not an eigenstate) propagating in the system. Let $u_i(t)$ denote the lattice displacement for coordinate i at time t. The total energy of the system can be written as a sum of local contributions:

$$E = \sum_i E_i \,, \qquad E_i = \frac{M_i}{2}\dot{u}_i^2 - \frac{1}{2}\sum_j u_i k_{ij} u_j \,. \tag{3.18}$$

Using $M_i\ddot{u}_i = -k_{ij}u_j$, the local change of energy with time is

$$\frac{\mathrm{d}E_i}{\mathrm{d}t} = \frac{1}{2}\sum_j (\dot{u}_i k_{ij} u_j - u_i k_{ij}\dot{u}_j) \equiv \sum_j J_{ij} \,. \tag{3.19}$$

The local energy current between each pair of local degrees of freedom is thus naturally defined as

$$J_{ij} = \frac{1}{2}(\dot{u}_i k_{ij} u_j - u_i k_{ij}\dot{u}_j) \,. \tag{3.20}$$

For a given phonon of frequency ω, we rewrite u and $\dot{u}$ in terms of the complex wave $\phi(t) \equiv \psi \mathrm{e}^{\mathrm{i}\omega t}$:

$$u_i(t) = \frac{\mathrm{Re}[\phi_i(t)]}{\sqrt{M_i}} \equiv \frac{\phi_i^{\mathrm{R}}(t)}{\sqrt{M_i}} \,, \qquad \dot{u}_i(t) = -\frac{\omega \mathrm{Im}[\phi_i(t)]}{\sqrt{M_i}} \equiv -\frac{\omega\phi_i^{\mathrm{I}}(t)}{\sqrt{M_i}} \,. \tag{3.21}$$

Hence, the current associated with that particular phonon between the i and j local degrees of freedom is

$$\begin{aligned} J_{ij} &= \frac{1}{2}\omega(\phi_i^{\mathrm{R}}\tilde{k}_{ij}\phi_j^{\mathrm{I}} - \phi_i^{\mathrm{I}}\tilde{k}_{ij}\phi_j^{\mathrm{R}}) \\ &= \frac{\omega}{4}(\phi_i^*\tilde{k}_{ij}\phi_j - \phi_i\tilde{k}_{ij}\phi_j^*) \,. \end{aligned} \tag{3.22}$$

The ϕ are solutions of the eigenvalue problem (3.3). The normalization condition for the phonon amplitude follows by equating the wave energy with $\hbar\omega$, whence

$$\sum_i |\phi_i|^2 = \frac{2\hbar}{\omega_\phi} \,. \tag{3.23}$$

Rather than use the phonon wave functions, it is more convenient to use wave functions normalized to 1: $\psi \equiv \sqrt{\omega_\phi/(2\hbar)}\phi$, such that $\sum_i |\psi_i|^2 = 1$. The current for a particular phonon mode is then obtained by adding the partial currents throughout the whole system cross-section. Using the matrix notation explained in Sect. 3.2.1, we can write this as

$$J_\psi = \frac{\hbar}{2}\left(\underline{\psi}_1^* \mathsf{k}_{1,2} \bar{\psi}_2 - \underline{\psi}_1 \mathsf{k}_{1,2} \bar{\psi}_2^*\right) \tag{3.24}$$

$$\equiv \frac{\hbar}{2}\mathscr{T}_\psi \,. \tag{3.25}$$

So we can already compute the current carried by any particular phonon mode. However, we are not done yet. We now have to add the contributions of all the phonons present in the system, weighted by their respective occupations. Since the system is not in equilibrium, the phonon occupation depends not only on its energy $\hbar\omega$, but also on whether it flows from the left into the right-hand reservoir, or vice versa. Phonons flowing from the left into the right-hand reservoir have an occupation corresponding to the Bose–Einstein distribution at the temperature of the left-hand reservoir:

$$N_\rightarrow = \frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T_\mathrm{l}} - 1} \,.$$

Those going from right to left have

$$N_\leftarrow = \frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T_\mathrm{r}} - 1} \,.$$

Therefore, we can sum the contributions from these two groups of phonons independently:

$$J = \int_0^\infty \left[N_\rightarrow(\omega) \sum_{\psi_\rightarrow} J_{\psi_\rightarrow} \delta(\omega^2 - \omega_{\psi_\rightarrow}^2) + N_\leftarrow(\omega) \sum_{\psi_\leftarrow} J_{\psi_\leftarrow} \delta(\omega^2 - \omega_{\psi_\leftarrow}^2) \right] \mathrm{d}(\omega^2) \,. \tag{3.26}$$

We have chosen ω^2 as our integration variable. We could have used ω instead, but it is convenient to write the δ function in terms of ω^2, so that we can later use property (3.6) of the resolvent. The previous equation can be simplified by noting that, for any eigenstate ψ, the state ψ^* is also an eigenstate propagating in the direction opposite to ψ. From (3.24), it is obvious that $J_\psi = -J_{\psi^*}$. Therefore we only need to sum over one of the two propagation directions, and (3.26) becomes

$$J = \frac{1}{2\pi} \int_0^\infty \hbar\omega \left[N_\rightarrow(\omega) - N_\leftarrow(\omega) \right] \mathscr{T}(\omega) \mathrm{d}\omega \,, \tag{3.27}$$

with

$$\mathscr{T}(\omega) \equiv 2\pi \sum_{\psi_\rightarrow} \mathscr{T}_{\psi_\rightarrow} \delta(\omega^2 - \omega_{\psi_\rightarrow}^2) \,, \tag{3.28}$$

where the $\mathscr{T}_\psi$ are defined in (3.25) and we have used $\mathrm{d}(\omega^2) = 2\omega\mathrm{d}\omega$. The function $\mathscr{T}(\omega)$ is a dimensionless quantity, usually called the transmission function. Once the transmission function has been computed, the thermal current at any temperatures of the two reservoirs is easily obtained. The thermal conductance σ is then

$$\sigma = \frac{1}{2\pi} \int_0^\infty \hbar\omega \frac{\mathrm{d}N}{\mathrm{d}T} \mathscr{T}(\omega) \mathrm{d}\omega \,. \tag{3.29}$$

So now our main task is to find ways of computing the transmission function for atomically defined systems.

3.2.3 Different Formulas for the Transmission

Using (3.28) to compute the transmission would require one to evaluate propagating wave functions in an infinitely extended system. This task is made possible by the use of the resolvent and the Lippmann–Schwinger equation. The resolvent formulation leads to the concept of transmission function in a natural way. Below we derive three common versions of this formulation: the three-region formula, the Landauer formula, and the two-region formula. The derivations presented here are perhaps not the most elegant, but they have the advantage of being quite direct and simple.

The Three-Region Formula

In this formulation, the system is divided into three regions. The sets of atoms at the two interfaces thus defined are labeled as 1, 2, 3, and 4 (see Fig. 3.3). These are not physical interfaces, but just arbitrarily placed artificial boundaries. The current transported by some state ψ can be evaluated at any cross-sectional interface, for example the one between sets 3 and 4. This is [see (3.24) and (3.25)]

$$\mathscr{T}_\psi = \underline{\psi}_3^* \mathsf{k}_{34} \overline{\psi}_4 - \underline{\psi}_3 \mathsf{k}_{34} \overline{\psi}_4^* \,. \tag{3.30}$$

Let us imagine for a moment that we artificially cut the connection between sets 1 and 2, and also between sets 3 and 4, by making $\mathsf{k}_{12} = 0$ and $\mathsf{k}_{34} = 0$. Let us denote the eigenstates corresponding to that decoupled system by ϕ. In such cases, there are no propagating states. One can form propagating eigenstates of the coupled system from the uncoupled system eigenstates, using the LS equation. For states propagating to the right, we use the ϕ of the left-hand decoupled system, so ϕ is non-zero only on the left of interface 1–2. In this way we guarantee that the form of the wave infinitely far from the junction, on the right-hand side, is exclusively that

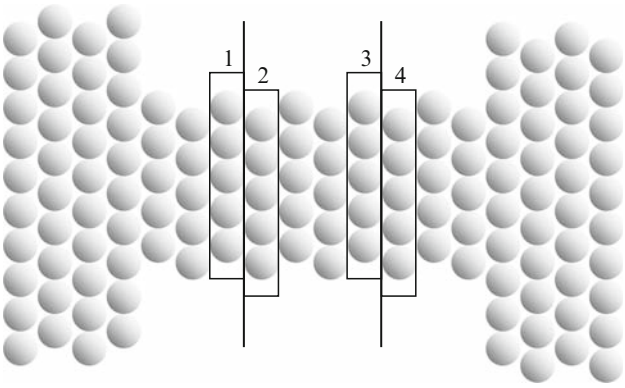

Fig. 3.3 Definition of the different atomic sets involved in the derivation of the three-region transmission formula

of a scattered wave. Therefore, at set 3, using (3.8), the wave function of the coupled system is

$$\overline{\psi}_3 = \mathsf{G}_{32}\mathsf{k}_{21}\overline{\phi}_1 \,, \tag{3.31}$$

where G is the resolvent of the coupled system. Let us denote the resolvent of the uncoupled system by g. Elements of g linking atoms on two different sides of an interface are zero. The wave function at set 4 is then

$$\overline{\psi}_4 = \mathsf{g}_{44}\mathsf{k}_{43}\overline{\psi}_3 = \mathsf{g}_{44}\mathsf{k}_{43}\mathsf{G}_{32}\mathsf{k}_{21}\overline{\phi}_1 \,. \tag{3.32}$$

By direct substitution into (3.30), we have

$$\mathscr{T}(\omega) = \mathrm{Tr}\sum_{\phi}\delta(\omega^2-\omega_\phi^2)\mathsf{G}_{32}\mathsf{k}_{21}\overline{\phi}_1\underline{\phi}_1^*\mathsf{k}_{12}[\mathsf{G}^+]_{23}\mathsf{k}_{34}(\mathsf{g}_{44}^+-\mathsf{g}_{44})\mathsf{k}_{43} \,.$$

Now, using the property of the local spectral density (3.6), we have

$$\mathscr{T}(\omega) = \mathrm{Tr}\big(\mathsf{G}_{32}\Gamma_2[\mathsf{G}^+]_{23}\Gamma_3\big) \,, \tag{3.33}$$

where $\Gamma_3 \equiv \mathsf{k}_{34}(\mathsf{g}_{44}-\mathsf{g}_{44}^+)\mathsf{k}_{43}$.

The Landauer Formula

Expression (3.33) can be written in a more symmetrical way [18]. For this, we define the matrix

$$\mathsf{t} = \Gamma_3^{1/2}\mathsf{G}_{32}\Gamma_2^{1/2} \,. \tag{3.34}$$

The existence of the matrix $\Gamma^{1/2}$ is guaranteed by the fact that Γ is positive definite. Then, using the cyclical property of the trace, the transmission becomes

$$\mathscr{T} = \mathrm{Tr}\big[\mathsf{t}\mathsf{t}^+\big] \,. \tag{3.35}$$

This is the Landauer form of the transmission formula. The matrix $\mathsf{t}\mathsf{t}^+$ is Hermitian, and its eigenvalues are bounded between 0 and 1 [18]. Each of these eigenvalues is associated with the transmission probability of one channel in the system, and their sum yields the total transmission.

The Two-Region Formula

There is an alternative formula that can also be used to compute the transmission. Unlike the previous ones, which involved three different subsets of atoms, this formula involves only two adjacent subsets. Therefore, the two-region formula is convenient when considering transport across an interface between two different materials, for example.

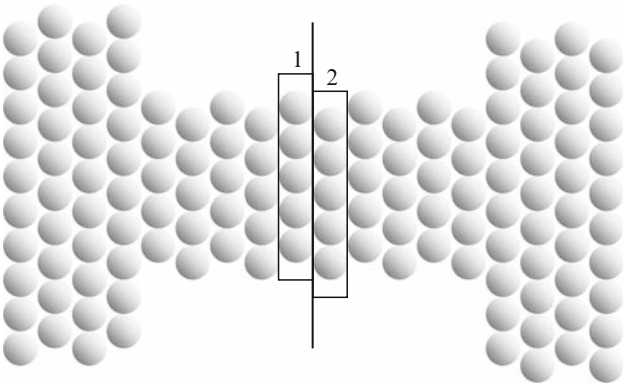

Fig. 3.4 Definition of the different atomic sets involved in the derivation of the two-region transmission formula

Let us label the two sets of atoms as 1 and 2 (see Fig. 3.4). By (3.24), the current carried by any particular phonon is

$$\mathscr{T}_\psi = \frac{1}{2}\left(\underline{\psi}_1 \mathsf{k}_{12} \overline{\psi}_2^* - \underline{\psi}_1^* \mathsf{k}_{12} \overline{\psi}_2\right) .$$

As in the last section, we used the Lippmann–Schwinger equation to construct the wave functions that propagate towards the right. As before, lower case symbols ϕ and g denote the wave function and Green function corresponding to the decoupled system, for which $\mathsf{k}_{12} = 0$. Symbols ψ and G, correspond to the coupled system. The Lippmann–Schwinger (LS) equation for the transposed wave function at region 1 states that

$$\underline{\psi}_1^* = \underline{\phi}_1^* + \underline{\psi}_1^* \mathsf{k}_{12} \mathsf{g}_2 \mathsf{k}_{21} \mathsf{g}_1 , \tag{3.36}$$

which implies

$$\underline{\psi}_1^* = \underline{\phi}_1^* \mathsf{D}_1^{\mathrm{A}} , \tag{3.37}$$

with

$$\mathsf{D}_1^{\mathrm{A}} \equiv \left(I - \mathsf{k}_{12} \mathsf{g}_2^+ \mathsf{k}_{21} \mathsf{g}_1^+\right)^{-1} . \tag{3.38}$$

As before, using the eigenfunctions ϕ of the left-hand uncoupled system guarantees that the resulting ψ will propagate to the right. Therefore $\phi_2 = 0$, and the LS equation in region 2 reads

$$\underline{\psi}_2^* = \underline{\psi}_1^* \mathsf{k}_{12} \mathsf{g}_2^+ . \tag{3.39}$$

Using (3.37), this becomes

$$\underline{\psi}_2^* = \underline{\phi}_1^* \mathsf{D}_1^{\mathrm{A}} \mathsf{k}_{12} \mathsf{g}_2^+ . \tag{3.40}$$

Multiplying the LS equation in region 1 on the left by k_{21}, we have

$$\mathsf{k}_{21} \psi_1 = \mathsf{k}_{21} \phi_1 + \mathsf{k}_{21} \mathsf{g}_1 \mathsf{k}_{12} \mathsf{g}_2 \mathsf{k}_{21} \psi_1 = \mathsf{D}_2^{\mathrm{R}} \mathsf{k}_{21} \phi_1 , \tag{3.41}$$

with $\mathsf{D}_2^{\mathrm{R}} \equiv (I - \mathsf{k}_{21} \mathsf{g}_1 \mathsf{k}_{12} \mathsf{g}_2)^{-1}$. Using the LS equation in region 2 as before, we have

$$\overline{\psi}_2 = \mathsf{g}_2\mathsf{k}_{21}\overline{\psi}_1 = \mathsf{g}_2\mathsf{D}_2^{\mathrm{R}}\mathsf{k}_{21}\overline{\phi}_1 \,. \tag{3.42}$$

Now we use (3.40) and (3.41) to express the first term of (3.24) in the form

$$\underline{\psi}_2^*\mathsf{k}_{21}\overline{\psi}_1 = \underline{\phi}_1^*\mathsf{D}_1^{\mathrm{A}}\mathsf{k}_{12}\mathsf{g}_2^+\mathsf{D}_2^{\mathrm{R}}\mathsf{k}_{21}\overline{\phi}_1 \,. \tag{3.43}$$

Using (3.37) and (3.42), the second term becomes

$$\underline{\psi}_1^*\mathsf{k}_{12}\overline{\psi}_2 = \underline{\phi}_1^*\mathsf{D}_1^{\mathrm{A}}\mathsf{k}_{12}\mathsf{g}_2\mathsf{D}_2^{\mathrm{R}}\mathsf{k}_{21}\overline{\phi}_1 \,. \tag{3.44}$$

The two expressions above are scalars. They can be transformed into an equivalent form, as the trace of a matrix. To do this, we move the $\overline{\phi}_1$ over to the left-hand side to form a matrix $\overline{\phi}_1\underline{\phi}_1^*$, and take the trace of the resulting matrix expression:

$$\underline{\psi}_2^*\mathsf{k}_{21}\overline{\psi}_1 - \underline{\psi}_1^*\mathsf{k}_{12}\overline{\psi}_2 = 2\pi\mathrm{Tr}\big(\overline{\phi}_1\underline{\phi}_1^*\mathsf{D}_1^{\mathrm{A}}\mathsf{k}_{12}\rho_2\mathsf{D}_2^{\mathrm{R}}\mathsf{k}_{21}\big) \,, \tag{3.45}$$

where $2\pi\rho = \mathsf{g} - \mathsf{g}^+$ is the spectral density of states of the decoupled system [see (3.6)]. Finally, recalling the definition of the total transmission in (3.28) and using (3.6), we get

$$\mathscr{T} = 4\pi^2\mathrm{Tr}\big(\rho_1\mathsf{D}_1^{\mathrm{A}}\mathsf{k}_{12}\rho_2\mathsf{D}_2^{\mathrm{R}}\mathsf{k}_{21}\big) \,. \tag{3.46}$$

This expression only involves the decoupled Green functions, g_1 and g_2, at the two regions defining the interface.

Problem 2. Compute the transmission of: (1) a perfect chain, (2) two linked equal chains, and (3) two chains with atoms of different mass.

Solution. Let us define the mass of the atoms on the left (labeled by integers) as $m = 1$. Then the mass of the atoms on the right is Δ. All the spring constants are k, except the one between atoms 1 and a, which has value t.

Let us use the two-region formula. We do this at the interface between atoms 1 and a. First we need to obtain the uncoupled Green functions. At atom 2 in Fig. 3.5, the uncoupled Green function (i.e., with $k_{21} = k_{12} = 0$) is the one obtained in Problem 1 [see (3.16)]:

$$g_{22}(z) = \gamma(z,k) \equiv \frac{z + 2k + \sqrt{z^2 + 4kz}}{2k^2} \,. \tag{3.47}$$

The uncoupled Green function at atom 1 is then

$$g_{11}(z) = \xi(z,k,t') \equiv \frac{1}{z + k + t' - k^2\gamma(z,k)} \,, \tag{3.48}$$

where $t' = t/\sqrt{\Delta}$. The uncoupled Green function at atom a is trivially

$$g_{aa}(z) = \xi(z,k/\Delta,t') \,. \tag{3.49}$$

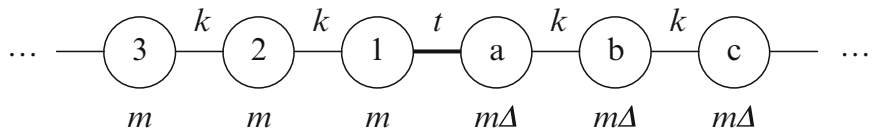

Fig. 3.5 System in Problem 2. Two linked semi-infinite one-dimensional chains, with different atomic masses, m and $m\Delta$. The interaction between the chains has a different spring constant (t) than the rest of the bonds (k)

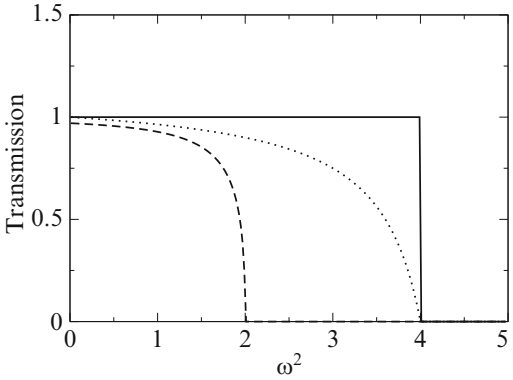

Fig. 3.6 Transmission function for the system depicted in Fig. 3.5. *Solid line*: $\Delta = 1$, $t = -1$. *Dotted line*: $\Delta = 1$, $t = -0.75$. *Dashed line*: $\Delta = 2$, $t = -1$

Thus, the transmission function is

$$\mathscr{T} = 4\frac{\mathrm{Im}[\xi(z,k,t')]t'^2\mathrm{Im}[\xi(z,k/\Delta,t')]}{\left|1-\xi(z,k,t')t'^2\xi(z,k/\Delta,t')\right|^2} . \tag{3.50}$$

Figure 3.6 shows particular results with $k = -1$, for the cases $\Delta = 1$, $t = -1$ (infinite chain), $\Delta = 1$, $t = -0.75$ (chain with one faulty link), and $\Delta = 2$, $t = -1$ (coupled semi-infinite chains with different masses).

3.2.4 Weak Coupling Limit: The Low Temperature Thermal Conductance of a Weak Junction

When two systems are only very weakly coupled, a simple, approximate expression can be obtained for the transmission function. Let us take the exact expression (3.46). Since we are in the weak coupling limit, we have $k_{12} \to 0$. Therefore $D_1 \simeq I$ and $D_2 \simeq I$, and we thus obtain the simple expression

$$\mathscr{T} \simeq 4\pi^2\mathrm{Tr}(\rho_1 k_{12}\rho_2 k_{21}) . \tag{3.51}$$

This is sometimes called the Bardeen approximation, because it was first used by J. Bardeen to treat the problem of electron tunneling [19]. This equation tells us that the weak coupling transmission function is proportional to the spectral densities of states of the contacts.

Let us now illustrate the application of this formula by deriving the form of the thermal conductance between two solids linked by a weak point contact. This problem was studied in [20]. Here we provide a simple derivation of the main result. First of all, we need to obtain the form of the spectral density of states of a solid surface in the low frequency limit. In Problem 1, we saw that the Green function at the 'surface' (in this case, the edge atom) of an atomic chain is [see (3.17)]

$$G_{11} = \frac{2}{\omega^2 - \sqrt{\omega^4 + 4k\omega^2}} \, .$$

The corresponding spectral density is therefore

$$\frac{1}{2\pi} \frac{\sqrt{\omega^4 + 4k\omega^2}}{2k\omega^2} \, ,$$

which, in the low frequency limit, goes as

$$\rho_{\mathrm{surf}}^{\mathrm{1D}} \propto \frac{1}{\omega} \, . \tag{3.52}$$

In the case of a real 3D surface, we can split the problem into multiple independent problems, each one corresponding to a different wave vector $\mathbf{q}_\parallel$ parallel to the surface. Because of parallel momentum conservation, each of these problems is completely decoupled from the rest, and can be described by an effective one-dimensional system. Now, instead of having $2k$ for the diagonal elements, we have to take into account the dispersion parallel to the surface, so the diagonal elements are $K_{ii} = 2k + c^2 q_\parallel^2$ at the inner layers, and $K_{11} = k + c^2 q_\parallel^2$ at the surface layer. Here c is the velocity of sound. So instead of (3.52), we have

$$\mathrm{Im}\big[G_{11}(q_\parallel)\big] \propto \frac{1}{\sqrt{\omega^2 - c^2 q_\parallel^2}} \, . \tag{3.53}$$

The total spectral density is obtained by integrating this over $\mathbf{q}_\parallel$:

$$\rho_{\mathrm{surf}}^{\mathrm{3D}} \propto \int_0^{\omega/c} \left(\omega^2 - c^2 q_\parallel^2\right)^{-1/2} q_\parallel \mathrm{d}q_\parallel \propto \omega \, . \tag{3.54}$$

Thus, the spectral density of a surface is proportional to ω. (In contrast, the bulk spectral density is proportional to ω^2, which is easily proven [21].) The transmission function is therefore proportional to the frequency squared: $\mathscr{T} \propto \rho^2 \propto \omega^2$. This implies that, for temperatures small compared with the Debye temperature, using (3.29), the thermal conductance goes as

$$\sigma \propto T^3 \, . \tag{3.55}$$

3.2.5 Upper Limits to Thermal Conductance, Entropy Flow, and Information Rates

Typically, anharmonic interactions reduce a system's thermal conductance compared with what it would be if it were harmonic. Similarly, impurities, defects, and disorder in general reduce the transmission function of a periodic system. Therefore, the transmission function of a perfect periodic system constitutes an upper bound for the transmission of the imperfect system, and it can be used to obtain theoretical bounds for several quantities, including the thermal conductance, the entropy flow, and the rate of information transfer through a system.

We illustrate this application in the case of transport through carbon nanotubes, as well as graphene and graphite. Experimental results confirm the validity of the resulting thermal conductance limits [9, 22, 23].

Maximum Thermal Conductance

The thermal conductance in the case of a perfect periodic system takes an appealingly simple form [24, 25]. We can write the thermal conductance in terms of the branch index α and wave vector k for the system:

$$\sigma^{\max}(T) = \sum_{\alpha} \int_0^{\pi/a} \frac{\partial \omega_{\alpha,k}}{\partial k} \hbar \omega_{\alpha,k} \frac{\mathrm{d}}{\mathrm{d}T} f(\omega_{\alpha,k}, T) \frac{\mathrm{d}k}{2\pi} , \tag{3.56}$$

and transform to an integral in frequency, to obtain (3.1), where now the transmission function takes only integer values:

$$\mathscr{T}^{\max}(\omega) \equiv \text{number of branches at frequency } \omega . \tag{3.57}$$

Single-Walled Carbon Nanotubes

Using (3.1) and (3.3), we can calculate upper bounds to the thermal conductivity of finite length nanotubes:

$$\kappa^{\max} = \frac{L}{s} \sigma^{\max} , \tag{3.58}$$

where L is the length of the suspended segment of the nanotube and s is its cross-section. (Note that $s = 2\pi R\delta$, where R is the nanotube radius and $\delta = 3.35$ Å the layer separation in graphite.)

To calculate the transmission functions, we have computed the nanotube phonon dispersion relations, using the interatomic potential of [26]. The reason for this

choice is that this potential gives a good description of the limiting case of a graphene sheet. This potential yields linear dispersions for the two lowest-lying flexural modes. It has been shown that these modes should have quadratic dispersions near the Γ point [27]. However, from (3.3) it is apparent that, to construct the transmission function, it does not matter whether the dispersions are linear or quadratic. What matters is that the upper and lower frequency limits of the branch should be accurately computed. In fact, we have confirmed that the potential of [27] yields very similar results to the ones presented here. The phonon dispersion relations shown in [26–28] were all reproduced, as a test of the implementation.

We show examples of transmission functions for a few zigzag nanotubes in Fig. 3.7. The transmission functions for the graphene and three-dimensional graphite limits are also shown, normalized by the cross-section. We see that as the diameter of the nanotube increases, its transmission function more closely resembles that of a graphene sheet. The computed upper bounds to the thermal conductance, divided by the nanotube cross-section, are shown in Fig. 3.8 as a function of temperature. At low T, the conductance of all nanotubes has a linear T dependence, with a prefactor $4\pi^2 k_B^2/3h$, corresponding to four quanta of thermal conductance. This is due to the transmission function being 4, independent of frequency, at the lowest frequencies. This is a general result for any type of wire, related to the four lowest-lying branches present in any free-standing wire [8, 25]. At higher temperatures, a T^2 dependence is achieved, as a result of higher phonon branches becoming active. As T increases, the curves converge to a single line, which finally saturates to a limiting high temperature value. Results for armchair tubes, not shown here, are virtually indistinguishable from those for zigzag tubes, when the same diameters are compared.

Graphene

The limiting case of a graphene sheet, given by the thick solid line, provides a lower bound for all the nanotube curves and converges with them in the high temperature limit. Unlike the nanotubes, graphene has a $T^{1.5}$ dependence at low T. This anomalous behavior also differs from the T^2 behavior expected for two-dimensional acoustic phonon gases. The reason is that one of the three acoustic branches in graphene has a quadratic rather than linear dependence on frequency. For this quadratic branch, the frequency depends on the wave vector as $\omega = \alpha q^2$. For a graphene stripe of width $D \to \infty$, the contribution of this branch to the transmission function at low frequency is

$$\mathscr{T}(\omega) \simeq \int_{-q(\omega)}^{q(\omega)} \frac{D}{2\pi} \mathrm{d}q = \frac{D}{\pi}\sqrt{\frac{\omega}{\alpha}}\,.$$

Similarly, the two linear branches contribute an amount $(D/\pi)(\omega/c^{a(b)})$, where $c^{a(b)}$ are the speeds of sound of the two linear acoustic branches. For $\omega \to 0$, their contribution is negligible compared to that of the quadratic branch. Therefore, the upper bound to the thermal conductance of graphene at low temperature goes as

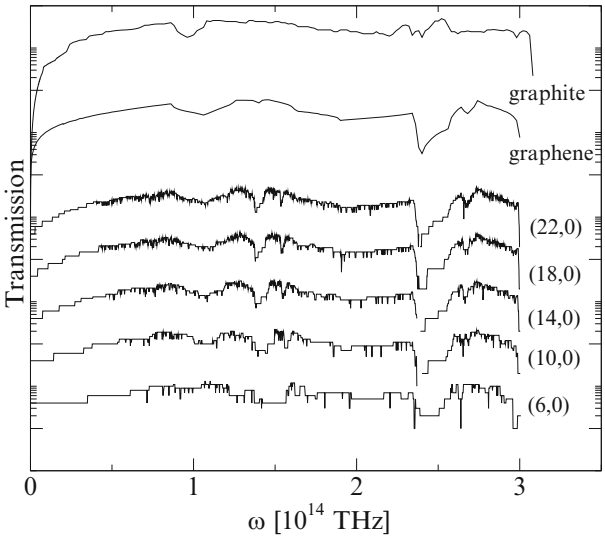

Fig. 3.7 Transmission functions for different diameters of single-walled carbon nanotubes (each rescaled by one order of magnitude for clarity), as well as graphene and graphite. The graphene and graphite results are scaled by the sample cross-section, and shown in arbitrary units

$$\frac{\sigma^{\max}}{s} = \frac{\eta(1/\delta)k_{\mathrm{B}}^{5/2}T^{3/2}}{2\pi^2\hbar^{3/2}\alpha^{1/2}}\,, \tag{3.59}$$

with

$$\eta \equiv \int_0^\infty \frac{x^{5/2}\mathrm{e}^x}{(\mathrm{e}^x-1)^2}\mathrm{d}x \simeq 4.46\,.$$

Substituting the value $\alpha = 0.62\times10^{-6}$ m^2/s, given by the theoretical dispersion, and also by other calculations [29], and using $\delta = 3.35$ Å, we obtain the result $\sigma^{\max}/s = 0.6\times10^6 T^{3/2}$ W/(m^{2}K$^{5/2}$).

Graphite

The thick dashed line in Fig. 3.8 shows the upper bound to the thermal conductance for three-dimensional graphite, in the basal plane, along the (110) direction. Accurate graphite dispersion relations were calculated using the method of [30]. At high T, the curve for graphite goes above the graphene and SWNT curves by about 20%. To understand this we note that the high temperature limit of σ is

$$\lim_{T\to\infty}\sigma(T) = \frac{k_{\mathrm{B}}}{2\pi}\int_0^\infty \mathscr{T}(\omega)\mathrm{d}\omega\,,$$

i.e., it is proportional to the area under the transmission curve. The frequency ranges of the phonon branches in graphite are generally larger than in graphene, due to the

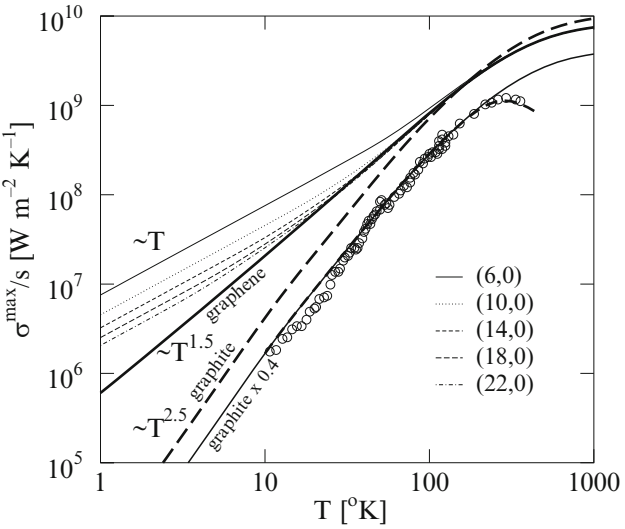

Fig. 3.8 Maximum thermal conductance divided by cross-section, for SWNTs, graphene, and graphite. Experimental results for MWNTs are proportional to the graphite curve, and only 0.4 times smaller

interlayer interaction. This results in a larger transmission function (at equal cross-sections) for graphite, and a somewhat larger high temperature upper bound.

At low temperature, the ballistic thermal conductance of graphite has a $T^{2.5}$ dependence. This limit can be obtained analytically, as for the graphene sheet. Now, there is some dispersion in the direction perpendicular to the planes, and the resulting transmission function in the low frequency limit is

$$\mathscr{T}_{\omega\to 0} = \frac{DW}{\pi^2\sqrt{\alpha}c}\frac{2}{3}\omega^{3/2}\times 2\,,$$

where DW is the cross-sectional area of the sample and c is the lowest speed of sound in the direction perpendicular to the planes. The factor of 2 arises from the double degeneracy of this branch. Repeating the argument of the previous paragraph, one obtains the $T^{5/2}$ dependence of the ballistic thermal conductance.

Maximum Entropy Flow and Information Rate

We have also calculated the maximum entropy and energy flow that can be carried through a carbon nanotube. This maximum flow would be attained in a system where a nanotube ballistically drains heat from a reservoir at temperature T into a reservoir at absolute zero. The energy and entropy flow are [24]

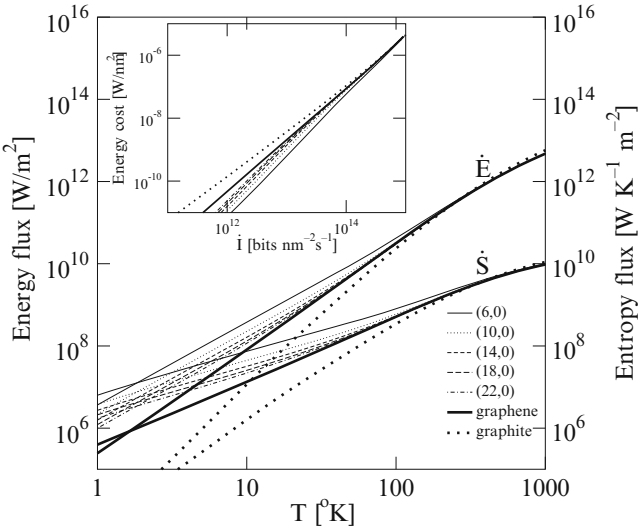

Fig. 3.9 Upper limits to energy and entropy flux, for SWNTs, graphene, and graphite, as a function of the temperature of the hot reservoir, when the cold reservoir is kept at 0 K. *Inset*: minimum energy expenditure necessary to mantain a given flux of information, for the same systems

$$\dot{E} = \int_0^\infty \mathscr{T}(\omega)\hbar\omega\left(\mathrm{e}^{\hbar\omega/k_\mathrm{B}T} - 1\right)^{-1}\frac{\mathrm{d}\omega}{2\pi}\,, \tag{3.60}$$

$$\dot{S} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi}\int_0^T \frac{\mathrm{d}T'}{T'}\frac{\mathrm{d}}{\mathrm{d}T'}\left(\frac{\hbar\omega}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T'} - 1}\right)\,. \tag{3.61}$$

These quantities, divided by the cross-sectional area, are shown in Fig. 3.9 for SWNTs, graphene, and graphite, as a function of temperature.

Using the analogy between information flow and entropy flow [24], $\dot{S}$ can be translated into units of bits/s, as $\dot{I} = \dot{S}/k_\mathrm{B}\log 2$. We can put this abstract concept into more physical terms. Let us say we have many bits of information written atomically on a surface which is kept at 0 K. Suppose we can erase bits by approaching them with one end of a carbon nanotube, the other end of which is connected to a 'hot' reservoir at temperature T, and allowing energy to flow from this reservoir along the nanotube onto the surface. Then $\dot{I}(T)$ corresponds to the maximum number of bits we can erase per unit time. By plotting $\dot{E}$ versus the corresponding $\dot{I}$ at that temperature, we can then determine the minimum amount of energy expenditure required to transmit information through the nanotube at a certain rate. This is plotted in the inset of Fig. 3.9.

Single-walled nanotubes follow the $\dot{E}^{\min} \propto \dot{I}^2$ relation predicted for transport through discrete channels [24]. As the information rate increases, the SWNT curves collapse into the curve for graphene. By inserting the asymptotic low frequency form of $\mathscr{T} \propto \omega^{1/2}$, appropriate for graphene, into (3.6) and (3.7), we obtain $\dot{E} \propto \dot{I}^{5/3}$. For graphite, the dependence is instead $\dot{E} \propto \dot{I}^{7/5}$, also slower than in the discrete channel case. At higher rates, all curves collapse into a single line which then

follows a $\dot{E} \propto \dot{I}^2$ dependence. At much higher temperatures, this regime evolves into an exponential one, although the nanotube would melt well before such high rates could be attained.

3.3 The Anharmonic Problem

The theory in the previous section dealt only with harmonic systems. Depending on the system, this can constitute an important limitation to its use. This section introduces an approach that allows one to include the effect of anharmonicity into the phonon transport. The theory presented here includes anharmonic interactions via a Keldysh diagrammatic approach. The Keldysh formalism [38–41] is a very powerful technique of theoretical physics that has been applied extensively to non-equilibrium electron problems in the past [42–48]. One advantage of the Keldysh approach is that it is not restricted to infinitesimal temperature differences, but can be applied to any finite temperature difference between the hot and cold bodies.

3.3.1 Many-Body Hamiltonian

The quantum mechanical lattice properties of any atomically described system can be represented by the Hamiltonian

$$\hat{H} = \frac{1}{2}\sum_{ij} k_{ij}\hat{\varphi}_i\hat{\varphi}_j + \frac{1}{2}\sum_{i} M_i\dot{\hat{\varphi}}_i\dot{\hat{\varphi}}_i + \sum_{ijk} V^{(3)}_{ijk}\hat{\varphi}_i\hat{\varphi}_j\hat{\varphi}_k + \cdots . \tag{3.62}$$

The terms on the right are the harmonic, kinetic and anharmonic terms respectively. $\hat{\varphi}_i$ is the Heisenberg displacement operator for the ith atomic degree of freedom. Here,

$$k_{ij} \equiv \frac{\partial^2 E}{\partial u_i \partial u_j} ,$$

$$V^{(3)}_{ijk} \equiv \frac{1}{3!}\frac{\partial^3 E}{\partial u_i \partial u_j \partial u_k} ,$$

E is the total energy, and the u_i are the atomic coordinates. M_i is the mass of the atom to which the ith degree of freedom belongs.

We can also write the Hamiltonian in the orthogonal representation, in terms of the eigenmodes of the harmonic part [12]:

$$\hat{H} = \sum_{q} \hbar\omega_q \left(\hat{b}^+_q \hat{b}_q + 1/2\right) + \sum_{ijk} V^{(3)}_{ijk}\hat{\varphi}_i\hat{\varphi}_j\hat{\varphi}_k + \cdots . \tag{3.63}$$

Here, $\hat{b}_q^+$ and $\hat{b}_q$ are the phonon creation and destruction operators, and the summation is over all the vibrational eigenstates q. The field operators are expressed in terms of the phonon creation and destruction operators as

$$\begin{cases} \hat{\varphi}_j(t) = \sum_q \sqrt{\hbar/M\omega_q}\left[\hat{b}_q^+ \phi_q^*(j)\mathrm{e}^{\mathrm{i}\omega_q t} + \hat{b}_q \phi_q(j)\mathrm{e}^{-\mathrm{i}\omega_q t}\right], \\ \hat{\dot{\varphi}}_j(t) = \sum_q \mathrm{i}\sqrt{\hbar\omega_q/M}\left[\hat{b}_q^+ \phi_q^*(j)\mathrm{e}^{\mathrm{i}\omega_q t} - \hat{b}_q \phi_q(j)\mathrm{e}^{-\mathrm{i}\omega_q t}\right], \end{cases} \tag{3.64}$$

where the eigenfunctions $\phi_q(j)$ satisfy

$$\sum_j k_{ij}\phi_q(j) = M_i\omega_q^2\phi_q(i)\,. \tag{3.65}$$

Inserting (3.64) into (3.62) and using (3.65), one obtains (3.63).

To simplify the following discussion, we will assume the same mass M for every atom in the system. Generalization to different masses is not difficult.

3.3.2 The Heat Current

Let us now derive an expression for the total heat current. In every non-equilibrium Green function problem, it is customary to divide the system into three parts: two contact reservoirs, plus a central system. Many-body interactions are only considered within the central system, but not at the contacts [42, 44]. In our case, this implies that the links joining the central system to the contacts are harmonic. We must note that the frontiers delimiting the central system can be arbitrarily defined, and there are no real interfaces separating it from the contacts. Therefore, one can place those harmonic links well within the reservoir, so that their lack of anharmonicity will not affect the total current, which is mostly limited by the narrowest part of the system, i.e., the actual junction or physical interface.

Let us consider the atoms at one of the contact frontiers, which are thus harmonic. For each degree of freedom within this subsystem, we can define a 'local energy' operator,

$$\hat{H}_i = \frac{1}{4}\sum_j (\hat{\varphi}_i k_{ij}\hat{\varphi}_j + \hat{\varphi}_j k_{ji}\hat{\varphi}_i) + \frac{M_i}{2}\hat{\dot{\varphi}}_i^2\,, \tag{3.66}$$

such that the total Hamiltonian of the subsystem is expressed as

$$\hat{H} = \sum_i \hat{H}_i\,. \tag{3.67}$$

The change in local energy with respect to time is then

$$\frac{\mathrm{d}E_i}{\mathrm{d}t} \equiv \frac{\mathrm{d}\langle \hat{H}_i \rangle}{\mathrm{d}t} = \left\langle \frac{1}{4}\sum_j \hat{\varphi}_i k_{ij} \hat{\varphi}_j + \hat{\varphi}_j k_{ji} \hat{\varphi}_i + \hat{\varphi}_i k_{ij} \hat{\varphi}_j + \hat{\varphi}_j k_{ji} \hat{\varphi}_i + \frac{M_i}{2}\left(\dot{\hat{\varphi}}_i \hat{\varphi}_i + \hat{\varphi}_i \dot{\hat{\varphi}}_i\right) \right\rangle , \tag{3.68}$$

which, using $M_i \dot{\hat{\varphi}}_i = -\sum_j k_{ij} \hat{\varphi}_j$, becomes

$$\frac{\mathrm{d}E_i}{\mathrm{d}t} = \frac{1}{4}\sum_j \left(\langle \dot{\hat{\varphi}}_i k_{ij} \hat{\varphi}_j \rangle - \langle \hat{\varphi}_i k_{ij} \dot{\hat{\varphi}}_j \rangle \right) - \left(\langle \dot{\hat{\varphi}}_j k_{ji} \hat{\varphi}_i \rangle - \langle \hat{\varphi}_j k_{ji} \dot{\hat{\varphi}}_i \rangle \right) \equiv \sum_j J_{ij} \, .$$

So we have obtained the expression for the local current between two different degrees of freedom belonging to two mutually interacting atoms. For a steady state, these local energy currents can be expressed in terms of phonon Green's functions as follows:

$$\begin{aligned} J_{ij} &= \frac{1}{4}\lim_{t\to t'} \frac{\mathrm{d}}{\mathrm{d}t}\left\{ k_{ji}\left[\langle \hat{\varphi}_j(t)\hat{\varphi}_i(t')\rangle - \langle \hat{\varphi}_j(t')\hat{\varphi}_i(t)\rangle\right] - k_{ij}\left[\langle \hat{\varphi}_i(t)\hat{\varphi}_j(t')\rangle - \langle \hat{\varphi}_i(t')\hat{\varphi}_j(t)\rangle\right]\right\} \\ &= \frac{\mathrm{i}\hbar}{4}\lim_{t\to t'} \frac{\mathrm{d}}{\mathrm{d}t}\left\{ k_{ij}\left[D^<_{ji}(t,t') - D^<_{ji}(t',t)\right] - k_{ji}\left[D^<_{ij}(t,t') - D^<_{ij}(t',t)\right]\right\} \\ &= \frac{1}{2}\int_{-\infty}^{\infty} \hbar\omega \left[k_{ij} D^<_{ji}(\omega) - D^<_{ij}(\omega) k_{ji}\right] \frac{\mathrm{d}\omega}{2\pi} \, . \end{aligned} \tag{3.69}$$

Here, a particular kind of time dependent and frequency dependent Green's function has made its appearance for the first time [40], namely

$$D^<_{ij}(t_1,t_2) \equiv -\frac{\mathrm{i}}{\hbar}\langle \hat{\varphi}_j(t_2)\hat{\varphi}_i(t_1)\rangle = \int_{-\infty}^{\infty} D^<_{ij}(\omega) \mathrm{e}^{-\mathrm{i}\omega(t_2-t_1)} \frac{\mathrm{d}\omega}{2\pi} \, .$$

Therefore, it is clear that one needs to compute the many-body Green's functions in order to obtain thermal currents. The way to obtain these Green's functions is explained below.

3.3.3 Computing the Interacting Phonon Green Functions

Now let us see how to compute the non-equilibrium phonon Green's functions for the type of system described above. There are four interrelated Green's functions [41]: $D^<$, $D^>$, D^{R}, and D^{A}. In the time representation they are defined as [12]

$$\mathrm{i}D^<_{lm}(t_1-t_2) \equiv \frac{1}{\hbar}\left\langle \hat{\varphi}_m(t_2)\hat{\varphi}_l(t_1)\right\rangle , \tag{3.70}$$

$$\mathrm{i}D^>_{lm}(t_1-t_2) \equiv \frac{1}{\hbar}\left\langle \hat{\varphi}_l(t_1)\hat{\varphi}_m(t_2)\right\rangle , \tag{3.71}$$

$$\mathrm{i}D^{\mathrm{R}}_{lm}(t_1-t_2) = \frac{1}{\hbar}\Theta(t_1-t_2)\left\langle [\hat{\varphi}_l(t_1), \hat{\varphi}_m(t_2)]\right\rangle , \tag{3.72}$$

$$\mathrm{i}D^{\mathrm{A}}_{lm}(t_1-t_2) = -\frac{1}{\hbar}\Theta(t_2-t_1)\left\langle[\hat{\varphi}_l(t_1),\hat{\varphi}_m(t_2)]\right\rangle . \tag{3.73}$$

The problem we want to solve is in a steady state. Therefore, we will not compute the time dependent Green's functions, but their Fourier transforms. In the non-interacting equilibrium case, such frequency dependent GFs can be expressed in terms of the resolvent, which is a frequency dependent matrix that can be readily computed from the force constant matrix. For the non-equilibrium case, however, one has to use 'kinetic equations', which involve both the equilibrium and non-equilibrium GFs. The next two sections detail how to obtain the equilibrium and non-equilibrium GFs for phonons, and how to use them in practical computations.

Calculating the Uncoupled Green Functions at the Contacts

Let us consider a case where the central system is decoupled from the contacts, so that all the force constants linking the central system to the contacts are equal to zero. Then no current is flowing and the systems are in equilibrium. We denote these uncoupled Green functions at the contacts by $D^{<}_0$, D^{R}_0. The other two GFs are related to these by [40] $D^{>}_{0,lj}(\omega) = D^{<}_{0,jl}(-\omega)$, $D^{\mathrm{A}}_{0,lj}(\omega) = [D^{\mathrm{R}}_{0,jl}(\omega)]^*$. The reason why we need to compute these uncoupled GFs is that they enter the expression for the contact self-energies, (3.88) and (3.89), as we shall see in the next section.

We can derive the expression for the unperturbed $D^{<}_0$ at the contacts as follows. First we obtain the time dependent Green's function. For this, we start from its definition (3.70). Using the eigenstate expansions of the field operators in (3.64), one obtains

$$\mathrm{i}D^{<}_{0\,lm}(t) = \sum_k \frac{1}{M\omega_k}\left[(N_{-k}+1)\mathrm{e}^{\mathrm{i}\omega_k t} + N_k \mathrm{e}^{-\mathrm{i}\omega_k t}\right]\phi_k(l)\phi^*_k(m) , \tag{3.74}$$

where N_k is the occupation of eigenstate k, and the summation is over all the eigenstates. Label k is a shorthand notation for a set of quantum numbers, such as the wave vector and the polarization. Label $-k$ denotes the same quantum numbers as k except those for the wave vector, which has the opposite direction.

We are, however, interested in the frequency dependent form of $D^{<}_0$, which we obtain by Fourier transforming the previous expression:

$$\mathrm{i}D^{<}_{0\,lm}(\omega) = \sum_k \frac{\pi}{M\omega_k}\left[(1+N_{-k})\delta(\omega+\omega_k) + N_k\delta(\omega-\omega_k)\right]\phi_k(l)\phi^*_k(m) . \tag{3.75}$$

However this expression in terms of the eigenfunctions is not practical for computational purposes. It is much more convenient to express the Green function in terms of the resolvent, $G(\omega^2) \equiv \left[(\omega+\mathrm{i}\delta)^2\mathsf{I}-\mathsf{K}\right]^{-1}$, where K is the force constant matrix of the system, defined by

$$k_{ij} \equiv \frac{1}{\sqrt{M_i M_j}}\frac{\partial^2 E}{\partial u_i \partial u_j} ,$$

and $\delta \to 0$. For this, we use the identity $\delta(x-y) = 2x\delta(x^2-y^2)\Theta(x)$, with $y>0$, and rewrite (3.75) in the form

$$\mathrm{i}D^<_{0\,lm}(\omega) = \sum_k \frac{\phi_k(l)\phi_k^*(m)}{M} 2\pi\delta(\omega^2-\omega_k^2)\Big[(1+N_{-k})\Theta(-\omega)+N_k\Theta(\omega)\Big] . \quad (3.76)$$

We now use the definition of the spectral density, viz.,

$$\rho_{lm}(\omega^2) \equiv \sum_k \phi_k(l)\phi_k^*(m)\delta(\omega^2-\omega_k^2) . \quad (3.77)$$

Substituting this into (3.76) yields, for the equilibrium case,

$$D^<_{0\,ij} = -\frac{2\pi\mathrm{i}}{M}\Big[\Theta(-\omega)\rho_{ij}(\omega^2)+N(\omega^2)\rho_{ij}(\omega^2)\Big] , \quad (3.78)$$

where $N(\omega^2)$ is the equilibrium phonon occupation at frequency ω. This expression is much more convenient for computational purposes, and it only requires one to compute ρ.

Now, the spectral density of states ρ is directly related to the resolvent G. We first note that

$$\begin{aligned} G_{lm}(\omega^2) &\equiv \sum_k \frac{\phi_k(l)\phi_k^*(m)}{(\omega+\mathrm{i}\delta)^2-\omega_k^2} \\ &= \sum_k \frac{1}{2\omega_k}\left(\frac{1}{\omega-\omega_k+\mathrm{i}\delta}-\frac{1}{\omega+\omega_k+\mathrm{i}\delta}\right) , \end{aligned} \quad (3.79)$$

which is trivially verified by substitution. The summation in k extends over all the eigenstates of the system. Taking the imaginary part of (3.79), one obtains the spectral density of states ρ:

$$\begin{aligned} \mathrm{Im}\big[G_{lm}(\omega^2)\big] &= \sum_k \frac{\pi\phi_k(l)\phi_k^*(m)}{2\omega_k}\big[\delta(\omega-\omega_k)-\delta(\omega+\omega_k)\big] \\ &= \sum_k \pi\phi_k(l)\phi_k^*(m)\delta(\omega^2-\omega_k^2) \\ &\equiv \pi\rho_{lm}(\omega^2) , \end{aligned} \quad (3.80)$$

where $\delta \to 0^+$.

Thus, to evaluate $D_0^<$, one only needs to compute the resolvent G and the Bose occupation factors N. These occupation factors are different for each of the contacts, because the latter are at two different temperatures.

To obtain the unperturbed retarded Green function we proceed similarly. First, the time-dependent form of the retarded Green function D^{R} is obtained directly from its definition in (3.71):

$$
\begin{aligned}
\mathrm{i}D^{\mathrm{R}}_{0\,lm}(t) &= \frac{1}{\hbar}\Theta(t)\left\langle\left[\hat{\varphi}_l(t),\hat{\varphi}_m(0)\right]\right\rangle \\
&= \Theta(t)\Bigg\{\sum_q \frac{\phi_q(l)\phi_q^*(m)}{2M\omega_q}\Big[N_q\mathrm{e}^{\mathrm{i}\omega_q t}+(1+N_{-q})\mathrm{e}^{-\mathrm{i}\omega_q t} \\
&\qquad\qquad -N_{-q}\mathrm{e}^{-\mathrm{i}\omega_q t}-(1+N_q)\mathrm{e}^{\mathrm{i}\omega_q t}\Big]\Bigg\} \\
&= \Theta(t)\sum_q \frac{\phi_q(l)\phi_q^*(m)}{2M\omega_q}\left(\mathrm{e}^{-\mathrm{i}\omega_q t}-\mathrm{e}^{\mathrm{i}\omega_q t}\right). \qquad (3.81)
\end{aligned}
$$

For the frequency dependent form, we Fourier transform the previous expression to obtain

$$
\begin{aligned}
D^{\mathrm{R}}_{0\,lm}(\omega) &= \sum_q \frac{\phi_q(l)\phi_q^*(m)}{2M\omega_q}\left(\frac{1}{\omega-\omega_q+\mathrm{i}\delta}-\frac{1}{\omega+\omega_q+\mathrm{i}\delta}\right) \\
&= \sum_q \frac{\phi_q(l)\phi_q^*(m)}{2M}\frac{2}{(\omega+\mathrm{i}\delta)^2-\omega_q^2}, \qquad (3.82)
\end{aligned}
$$

so finally,

$$
D^{\mathrm{R}}_{0\,lm}(\omega) = \frac{1}{M}G_{lm}(\omega^2). \qquad (3.83)
$$

The Coupled Green Functions

If we couple the system, a net phonon current flows from the hotter to the colder contact, and we no longer have an equilibrium system. The GFs for the coupled system satisfy the Dyson equation [40], and from it, kinetic equations are derived for the particular GFs. This is a result of the general theory of non-equilibrium GFs. A common way of writing these equations is [41]:

$$
\mathsf{D}^{<(>)} = \mathsf{D}^{\mathrm{R}}\Sigma^{<(>)}\mathsf{D}^{\mathrm{A}}, \qquad (3.84)
$$

$$
\mathsf{D}^{\mathrm{R}} = \left[\omega^2\mathsf{I}-\mathsf{K}-\Sigma^{\mathrm{R}}\right]^{-1}, \qquad (3.85)
$$

$$
\mathsf{D}^{\mathrm{A}} = [\mathsf{D}^{\mathrm{R}}]^{+}. \qquad (3.86)
$$

The Σ are self-energy matrices [40]. They consist of a contribution $\Sigma_{\mathrm{h(c)}}$ due to the contact leads, plus a many-body contribution Σ_{M} from the anharmonic interactions within the island:

$$
\Sigma = \Sigma_{\mathrm{M}}+\Sigma_{\mathrm{h}}+\Sigma_{\mathrm{c}}. \qquad (3.87)
$$

It can be shown [41] that the contributions from the contacts are

$$\Sigma^{<}_{\mathrm{h(c)}} = \mathsf{K}_{\mathrm{h(c)}} \mathsf{D}^{<}_{0\,\mathrm{h(c)}} \mathsf{K}^{T}_{\mathrm{h(c)}} \,, \tag{3.88}$$

$$\Sigma^{\mathrm{R}}_{\mathrm{h(c)}} = \mathsf{K}_{\mathrm{h(c)}} \mathsf{D}^{\mathrm{R}}_{0\,\mathrm{h(c)}} \mathsf{K}^{T}_{\mathrm{h(c)}} \,, \tag{3.89}$$

where $\mathsf{K}_{\mathrm{h(c)}}$ is the part of the force constant matrix joining the central system to the hot (cold) contact.

The Many-Body Self-Energies

The many-body contributions to the self-energy are computed from the corresponding Feynman diagrams, following the general rules in [40]. The 3-phonon processes are represented by the diagrams in Fig. 3.10a and b, corresponding to the self-energies (3.90) and (3.91), respectively:

$$\mathrm{i}\Sigma^{<}_{\mathrm{M}(3)i,n}(\omega) = \hbar \sum_{jklm} \int_{-\infty}^{\infty} V^{(3)}_{ijk} D^{<}_{jl}(\omega') D^{<}_{km}(\omega - \omega') V^{(3)}_{lmn} \mathrm{d}\omega' \tag{3.90}$$

$$\mathrm{i}\Sigma^{<}_{\mathrm{M}(3)i,n}(\omega) = \hbar \sum_{jklm} \int_{-\infty}^{\infty} V^{(3)}_{ijk} D^{>}_{jl}(\omega') D^{<}_{km}(\omega + \omega') V^{(3)}_{lmn} \mathrm{d}\omega' \,. \tag{3.91}$$

Since $D^{>}_{lj}(\omega) = D^{<}_{jl}(-\omega)$ [40], the two self-energies above are equivalent. This 3-phonon self-energy is purely imaginary. Once $\Sigma^{<(>)}_{\mathrm{M}(3)}$ have been computed, the $\Sigma^{\mathrm{R(A)}}_{\mathrm{M}(3)}$ can be obtained from them. The imaginary part is given by [40]

$$2\mathrm{Im}\Sigma^{\mathrm{R}}_{\mathrm{M}(3)} = \Sigma^{<}_{\mathrm{M}(3)} - \Sigma^{>}_{\mathrm{M}(3)} \,. \tag{3.92}$$

The real part is related to the imaginary part by the Hilbert transform [43]:

$$\mathrm{Re}\Sigma^{\mathrm{R}}_{\mathrm{M}(3)} = \mathscr{H}\left[\mathrm{Im}\Sigma^{\mathrm{R}}_{\mathrm{M}(3)}\right] \,. \tag{3.93}$$

This Hilbert transform is evaluated numerically, by the standard method of convoluting in the Fourier space.

The lowest order 4-phonon diagram is shown in Fig. 3.10c. This diagram does not contribute to $\Sigma^{<(>)}_{\mathrm{M}}$. However, it affects $\Sigma^{\mathrm{R(A)}}_{\mathrm{M}}$:

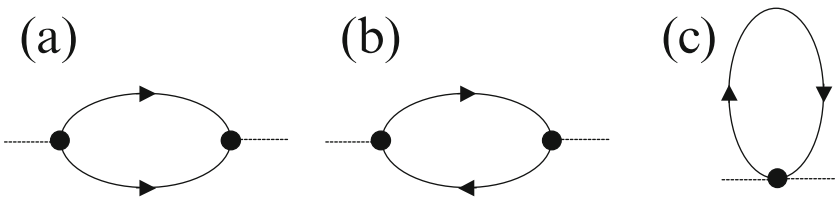

Fig. 3.10 Lowest order diagrams for the phonon many-body interactions

$$\mathrm{i}\Sigma^{\mathrm{R}}_{\mathrm{M}(4)i,n} = \hbar \sum_{jk} \int_{-\infty}^{\infty} V^{(4)}_{iklj} \left[D^{<}_{kl}(\omega') + D^{>}_{lk}(\omega') \right] \mathrm{d}\omega' \,. \tag{3.94}$$

Unlike the 3-phonon contribution, this 4-phonon contribution is purely real, and independent of frequency. This means that the fourth order diagrams in Fig. 3.10c introduce an elastic scattering of the phonons, analogous to the Hartree term for electrons.

In the example shown in the next section, we will only include 3-phonon processes for simplicity.

Self-Consistent Procedure

The many-body self-energies involve the total Green's functions. For this reason, in the presence of many-body interactions, the calculation needs to be done self-consistently. The complete calculation procedure is as follows. First, the resolvent at the contacts is computed, using projection techniques like those in [16] or [14], for example. The unperturbed Green's functions at the contacts are then obtained via (3.78) and (3.83). The contribution of the contacts to the self-energy is computed from these Green's functions via (3.88) and (3.89). For the first iteration, the many-body self-energies are taken to be zero. Then the self-consistent loop starts, by computing successively $\mathsf{D}^{\mathrm{R(A)}}$ [using (3.85) and (3.86)], $\mathsf{D}^{>(<)}$ [using (3.84)], $\Sigma^{<(>)}$ [using (3.87), (3.90), and (3.91)], and $\Sigma^{\mathrm{R(A)}}$ [using (3.92) and (3.93)]. This is repeated until convergence is achieved. The heat current can then be computed for the self-consistent system.

3.3.4 Another Formula for the Heat Current

Although (3.69) already expresses the current in terms of Green functions, we can obtain a nicer expression by following the steps given in [41]. To do this, we first define the 'current density' matrix, composed of elements that appear in (3.69):

$$\mathsf{J}(\omega) = \hbar\omega \left(\mathsf{K}\mathsf{D}^{<} - \mathsf{D}^{<}\mathsf{K} \right) \,. \tag{3.95}$$

From (3.69), it is obvious that the trace of this matrix, integrated over frequency, gives the total current coming in or going out of the system. We now use the following equations, which are straightforwardly obtained from the Dyson equation for the retarded or advanced Green function:

$$\mathsf{K}\mathsf{D}^{\mathrm{R}} = \omega^2 \mathsf{D}^{\mathrm{R}} - \Sigma^{\mathrm{R}}\mathsf{D}^{\mathrm{R}} - \mathsf{I} \,, \tag{3.96}$$

$$\mathsf{D}^{\mathrm{A}}\mathsf{K} = \omega^2 \mathsf{D}^{\mathrm{A}} - \mathsf{D}^{\mathrm{A}}\Sigma^{\mathrm{A}} - \mathsf{I} \,. \tag{3.97}$$

Substituting (3.84) into (3.95) and using (3.96) and (3.97), we have

$$\begin{aligned} \mathsf{K}\mathsf{D}^{<} - \mathsf{D}^{<}\mathsf{K} &= \mathsf{D}^{\mathrm{R}}\Sigma^{<}\mathsf{D}^{\mathrm{A}}(\Sigma^{\mathrm{A}} - \Sigma^{\mathrm{R}}) - \Sigma^{<}\mathsf{D}^{\mathrm{A}} + \mathsf{D}^{\mathrm{R}}\Sigma^{<} \\ &= \mathsf{D}^{<}(\Sigma^{\mathrm{A}} - \Sigma^{\mathrm{R}}) - \Sigma^{<}(\mathsf{D}^{\mathrm{A}} - \mathsf{D}^{\mathrm{R}}) \\ &= \mathsf{D}^{<}\Sigma^{>} - \Sigma^{<}\mathsf{D}^{>} , \end{aligned} \tag{3.98}$$

where we have used

$$\Sigma^{\mathrm{A}} - \Sigma^{\mathrm{R}} = \Sigma^{>} - \Sigma^{<} , \qquad \mathsf{D}^{\mathrm{A}} - \mathsf{D}^{\mathrm{R}} = \mathsf{D}^{>} - \mathsf{D}^{<} . \tag{3.99}$$

Therefore, the current frequency distribution at the hot (h) or cold (c) contact interfaces is

$$J_{\mathrm{h(c)}}(\omega) = +(-)\mathrm{Tr}\left[\Sigma^{<}_{\mathrm{h(c)}}\mathsf{D}^{>} - \Sigma^{>}_{\mathrm{h(c)}}\mathsf{D}^{<}\right]\frac{\hbar\omega}{2\pi} . \tag{3.100}$$

Integrated over frequency, this yields the total heat current. In general, if there are inelastic processes, $J_{\mathrm{h}}(\omega) \neq J_{\mathrm{c}}(\omega)$, but their integrals are the same.

3.3.5 Can We 'See' the Phonon Current?

A valuable feature of the real space description is that it allows one to visualize the local energy current density and track how inelastic processes spread the phonon frequency distribution. Expressing the current in terms of (3.100), one can visualize how phonons are scattered between frequencies when they cross the system. Let us illustrate this for the case of an infinitely long one-dimensional chain of atoms containing just one anharmonic link (see inset of Fig. 3.11). The spring constants and masses are taken to be unity, so the allowed frequency range is between 0 and 2. We show a case of very strong 3-phonon anharmonicity, with $V^2_{(3)} = M^{1/2}k^{5/2}/\hbar$. We inject a uniform distribution of phonons in a window of frequencies, with an

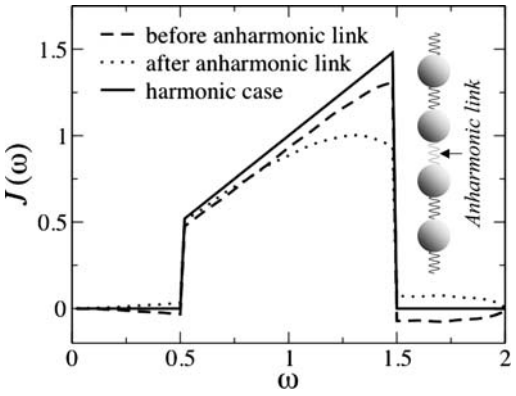

Fig. 3.11 Local energy current distribution for a window jet of phonons in a linear chain containing an anharmonic link, evaluated at two different points in the chain

occupation factor

$$N(\omega) = \begin{cases} 1\,, & \text{if } 0.5 < \omega < 1.5\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{3.101}$$

The current frequency distribution at the left and right contacts with the anharmonic link is shown in Fig. 3.11. If the link is harmonic, transport is elastic, and the current distribution extends over the same range as the impinging jet of phonons. In that case, the current distributions before and after the central link are exactly the same. On the other hand, when the link is anharmonic, the current distribution broadens with respect to that of the impinging phonon jet. Moreover, the current distributions are different before and after the link. This is because phonons interact inelastically. The distribution before the link is negative at frequencies out of range of the phonon jet, corresponding to inelastically reflected phonons. Past the link, the distribution is positive at all frequencies, corresponding to phonons exiting the system on that side. Equation (3.100) thus provides a way of visualizing and interpreting the inelastic flow of phonons through the system.

3.4 Concluding Remarks

Phonon transport has been formulated with Green's functions in a number of publications. It is not our goal to account for everything that has been published, but only to comment in general on what has been done, and what problems need to be addressed.

The Green's function approach to phonon transmission in harmonic systems has been presented in different work based on an elastic continuum model [52,53]. With this approach, it has been possible to address the effect of surface disorder on the thermal conductance of nanowires. Moreover, the effect of scattering at T-shaped structures has been investigated. The continuum formulation is very useful for large structures at low temperature, when the phonon mean free path is comparable to the size of the structure. One limitation of the continuum approach is that its phonon spectrum is not bounded on the upper frequencies. The thermal conductance would thus continue to increase as the temperature rises, unless the spectrum were cut off.

Atomistic Green's function formulations of the phonon transmission across harmonic structures have been presented in several studies [17, 37, 54–56]. One of these focused on conduction along one-dimensional chains linking two large reservoirs [55]. Another studied the thermal conductance of disordered solid–solid interfaces [56]. In the latter, isotopic mass disorder was considered on an FCC lattice with central potential interactions. An atomistic, three-dimensional formulation was also used to study phonon transport across Si nanowires, partially coated with an amorphous material [37]. The potential in this case was a variant of the Stillinger–Weber type, and went beyond the central potential model by including also 3-body terms. The case of transport at Si–Ge interfaces, where the interface

is fully connected, has been addressed in [17]. Different authors have also studied harmonic phonon transport at the nanoscale using various approaches [57–60].

In general, each of the above aimed to elucidate some particular physical behavior, so the models used are only realistic to a certain extent. From the standpoint of computational materials science, however, it is important to move towards increasingly realistic models. One important feature is the use of realistic interatomic potentials. There are a number of potentials in the literature for particular materials. Problems arise, however, when these potentials are used in situations when the coordination of the atoms changes, or when multiple species are involved. But these are precisely the kind of systems that one studies with Green's functions. If no adequate interatomic potentials can be devised for a particular structure, there is the alternative of using first-principles total energy calculations. This is feasible if the problem is mostly harmonic. Then the force constants can be calculated using ab initio techniques [13]. In some cases, these force constants may extend to a very large number of neighbors [61]. Thus the problem of computing the Green's functions will clearly be harder than if only short-range interactions are involved, because the size of the matrices to be inverted is directly related to the range of the potential. It is important to study the way different results are obtained depending on the range of the potential, and to address the question of accuracy given by different ab initio techniques [13].

One problem with atomistic methods is their size limitation. In order to be able to study large systems, real space Green's functions techniques must be used [62]. In certain cases these techniques can reduce the computation time to a linear scaling with system size.

Nevertheless, it would be unrealistic to think that the size of atomistic calculations might be enlarged so much as to displace coarser-grained techniques such as the Boltzmann transport equation (BTE). A very urgent need is thus to interface GF and BTE techniques, so that the output from the former can be used as an input for the latter. An obvious place where this is needed is in imposing interfacial boundary conditions for the BTE based on atomistic calculations of phonon transmission and reflection at the interface. This may be done either with a Green's function or a lattice dynamical approach, which should be equivalent, since the physics involved is the same.

The idea of linking GF with the BTE in a coarse-graining type of model is important for the study of thermal conductivity in nanocomposites, where many transport regimes and characteristic lengths are involved. On the one hand, diffusive transport can take place through extended regions of the matrix. On the other hand, ballistic transport may occur across nanoparticles, and the atomic character of their connexion to the matrix may be a determining factor in the way they modify the heat current.

It is equally important to compare GF and molecular dynamics calculations of thermal transport. Some steps have been taken, by computing transmission coefficients of wave packets using molecular dynamics [63]. By comparing these two techniques, it will be possible to determine the relative importance of anharmonic scattering, and of quantum mechanical effects, on the thermal conductance of the system.

Acknowledgements I wish to thank Prof. D.A. Broido from Boston College for his collaboration in several results shown in Sect. 3.2.5, and for invaluable advice throughout this work. I am very grateful to Dr. Wei Zhang and Prof. T.S. Fisher, from Purdue University, for fruitful and enthusiastic scientific discussions.

References

1. A. Majumdar: J. Heat Transfer **115**, 7 (1993)
2. R. Prasher: Appl. Phys. Lett. **83**, 48 (2003)
3. A.J.C. Ladd, B. Moran, and W.G. Hoover: Phys. Rev. B **34**, 5058 (1986)
4. P.B. Allen and J.L. Feldman: Phys. Rev. B **48**, 12581 (1993)
5. D.A. Young and H.J. Maris: Phys. Rev. B **40**, 3685 (1989)
6. S. Pettersson and G.D. Mahan: Phys. Rev. B **42**, 7386–7390 (1990)
7. J. Wang and J.S. Wang: `cond-mat/0509092`
8. K. Schwab, E.A. Henriksen, J.M. Worlock, and M.L. Roukes: Nature **404**, 974 (2000)
9. P. Kim, L. Shi, A. Majumdar, and P.L. McEuen: Phys. Rev. Lett. **87**, 215502 (2001)
10. D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar: Appl. Phys. Lett. **83**, 2934 (2003)
11. P.K. Schelling, S.R. Phillpot, and P. Keblinski: Phys. Rev. B **65**, 144306 (2002)
12. E.N. Economou: *Green's Functions in Quantum Physics*, Springer-Verlag, Berlin (1983)
13. N. Mingo, D.A. Stewart, D.A. Broido, and D. Srivastava: Phys. Rev. B **77**, 033418 (2008)
14. F. Guinea, C. Tejedor, F. Flores, and E. Louis: Phys. Rev. B **28**, 4397 (1983)
15. M. Lannoo and P. Friedel: *Atomic and Electronic Structure of Surfaces*, Springer (1991)
16. R. Haydock, V. Heine, and M.J. Kelly: J. Phys. C: Solid State Phys. **5**, 2845 (1972)
17. W. Zhang, N. Mingo, and T.S. Fisher: Journal of Heat Transfer (accepted, 2006)
18. J.C. Cuevas, A. Levy Yeyati, and A. Martin-Rodero: Phys. Rev. Lett. **80**, 1066 (1998)
19. J. Bardeen: Phys. Rev. Lett. **57**, 6 (1961)
20. M.P. Blencowe: Phys. Rev. B **59**, 4992 (1999)
21. G.P. Srivastava: *The Physics of Phonons*, Adam Hilger, Bristol (1990)
22. C. Yu, L. Shi, Z. Yao, D. Li, A. Majumdar: Nano. Lett. **5**, 1842–1846 (2005)
23. E. Pop, D. Mann, Q. Wang, K.E. Goodson, and H. Dai: Nano Letters **6**, 96 (2006)
24. J.B. Pendry: J. Phys. A **16**, 2161 (1983)
25. L.G.C. Rego and G. Kirczenow: Phys. Rev. Lett. **81**, 232 (1998)
26. R. Saito, G. Dresselhaus, and M.S. Dresselhaus: *Physical Properties of Carbon Nanotubes*, Imperial College Press, London (1998)
27. G.D. Mahan and G.S. Jeon: Phys. Rev. B **70**, 075405 (2004)
28. J. Maultzsch, S. Reich, C. Thomsen, E. Dobardžić, I. Milošević, and M. Damjanović: Sol. State Commun. **121**, 471 (2002)
29. D. Sánchez-Portal, E. Artacho, J.M. Soler, A. Rubio, and P. Ordejón: Phys. Rev. B **59**, 12678 (1999)
30. R.A. Jishi, L. Venkataraman, M.S. Dresselhaus, and G. Dresselhaus: Chem. Phys. Lett. **209**, 77 (1993)
31. N. Mingo and D.A. Broido: Nano Lett. **5**, 1221–1225 (2005)
32. M.G. Holland, C.A. Klein, and W.D. Straub: J. Phys. Chem. Solids **2**7, 903 (1966)
33. B.T. Kelly: *Physics of Graphite*, Applied Science Publishers, London (1981)
34. C.A. Klein and M.G. Holland: Phys. Rev. **136**, A575 (1964)
35. For [32], the crystallite length can be estimated to be about 5 micrometers
36. S. Berber, Y-K. Kwon, and D. Tománek: Phys. Rev. Lett. **84**, 4613 (2000)
37. N. Mingo and L. Yang: Phys. Rev. B **68**, 245406 (2003)
38. L.V. Keldysh: Sov. Phys. JETP **20**, 1018 (1965)
39. K.T. Mahanthappa: Phys. Rev. **126**, 329 (1961); P.M. Baksi and K.T. Mahanthappa: J. Math. Phys. **4**, 1 (1963); Ibid. **4**, 12 (1963)

40. E.M. Lifshitz and L.P. Pitaevskii: *Physical Kinetics*, Course of Theoretical Physics, Vol. 10, Pergamon Press, Oxford (1981)
41. S. Datta: *Electronic Transport in Mesoscopic Systems*, Cambridge University Press (1995)
42. R. Combescot: J. Phys. C **4**, 2611 (1971)
43. E.V. Anda and F. Flores: J. Phys. Cond. Matter **2**, 8023 (1991)
44. Y. Meir and N.S. Wingreen: Phys. Rev. Lett. **68**, 2512 (1992)
45. K. Makoshi and T. Mii: Surf. Sci. **357–358**, 335 (1995)
46. L.E. Henrickson: J. Appl. Phys. **91**, 6273 (2002)
47. D.A. Stewart and F. Leonard: Phys. Rev. Lett. **93**, 107401 (2004)
48. M. Galperin and A. Nitzan: Phys. Rev. Lett. **95**, 206802 (2005)
49. J.E. Lennard-Jones: Proceedings of the Physical Society **43**, 461–482 (1931)
50. F. Stillinger and T.A. Weber: Phys. Rev. B **31**, 5262 (1985)
51. N.W. Ashcroft and N.D. Mermin: *Solid State Physics*, Harcourt (1976)
52. A. Kambili, G. Fagas, V.I. Fal'ko, and C.J. Lambert: Phys. Rev. B **60**, 15593 (1999)
53. D.H. Santamore and M.C. Cross: Phys. Rev. B **63**, 184306 (2001); D.H. Santamore and M.C. Cross: Phys. Rev. B **66**, 144302 (2002); D.H. Santamore and M.C. Cross: Phys. Rev. Lett. **87**, 115502 (2001); M.C. Cross and R. Lifshitz: Phys. Rev. B **64**, 085324 (2001)
54. D.E. Angelescu, M.C. Cross, and M.L. Roukes: Superlattices and Microstruct. **23**, 673 (1998)
55. A. Ozpineci and S. Ciraci: Phys. Rev. B **6**3, 125415 (2001); A. Buldum, S. Ciraci, and C.Y. Fong: J. Phys.: Condens. Matter **12**, 3349 (2000); S. Ciraci, A. Buldum, and I.P. Batra: ibid. **13**, R537 (2001)
56. G. Fagas, A.G. Kozorezov, C.J. Lambert, and J.K. Wigmore: Phys. Rev. B **60**, 6459 (1999)
57. D. Segal, A. Nitzan, and P. Hänggi: J. Chem. Phys. **119**, 6840 (2003)
58. D.M. Leitner and P.G. Wolynes: Phys. Rev. E **61**, 2902 (2000)
59. K.R. Patton and M.R. Geller: Phys. Rev. B, **64** 155320 (2001)
60. C.-M. Chang and M.R. Geller: Phys. Rev. B **71**, 125304 (2005)
61. O. Dubay and G. Kresse: Phys. Rev. B **67**, 035401 (2003)
62. S.Y. Wu, J. Cocks, and C.S. Jayanthi: Phys. Rev. B **49**, 7957 (1994)
63. P.K. Schelling, S.R. Phillpot, and P. Keblinski: Appl. Phys. Lett. **80**, 2484 (2002)

Chapter 4
Macroscopic Conduction Models by Volume Averaging for Two-Phase Systems

Benoît Goyeau

The aim here is to describe macroscopic models of conductive heat transfer within systems comprising two solid phases, using the method of volume averaging. The presentation of this technique largely stems from work by Carbonell, Quintard, and Whitaker [1–3]. The macroscopic conservation equations are set up under the assumption of local thermal equilibrium, leading to a model governed by a single equation. The effective thermal conductivity of the equivalent medium is obtained by solving the associated closure problems. The case where thermal equilibrium does not pertain, leading to a model with two energy conservation equations, is discussed briefly.

4.1 Introduction

We consider heat conduction in a composite system combining two phases. In general, this configuration may correspond to a porous medium saturated by an immobilised phase, but also a composite medium comprising two solid phases. It is mainly the latter case that will be discussed in the following. For simplicity, the two solid phases will be assumed to be non-deformable, with constant thermophysical properties. Under such conditions, the local energy conservation equations within each of the phases σ and β (see Fig. 4.1) are

$$(\rho c_p)_\sigma \frac{\partial T_\sigma}{\partial t} = \nabla\cdot(k_\sigma \nabla T_\sigma) \quad \text{in phase } \sigma\,, \tag{4.1}$$

and

$$(\rho c_p)_\beta \frac{\partial T_\beta}{\partial t} = \nabla\cdot\left(k_\beta \nabla T_\beta\right) \quad \text{in phase } \beta\,. \tag{4.2}$$

The boundary conditions at the fluid–solid interface $A_{\beta\sigma}$, expressing continuity of temperature and heat flux, can be written in the form

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_4, © Springer-Verlag Berlin Heidelberg 2009

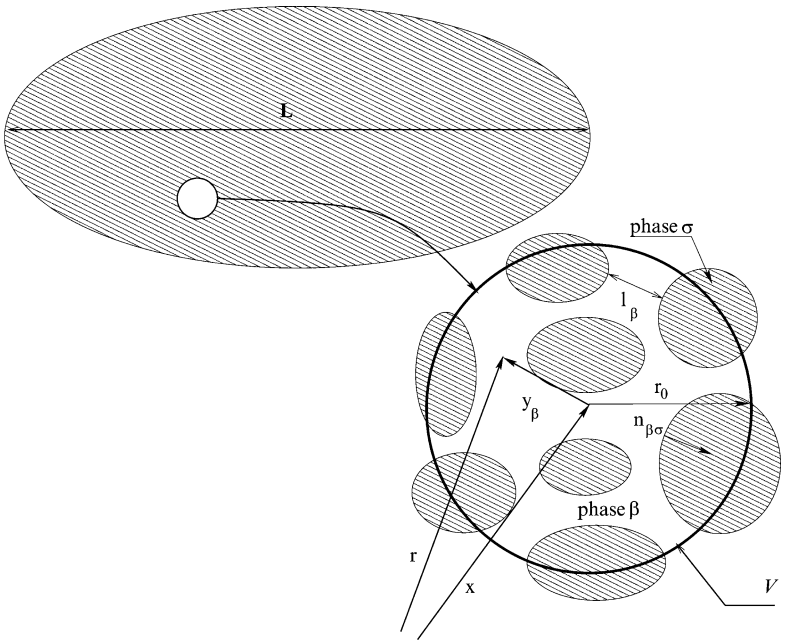

Fig. 4.1 Volume averaging region V for a system comprising two phases σ and β

$$T_\beta = T_\sigma \,, \quad \text{on } A_{\beta\sigma} \,, \tag{4.3}$$

and

$$\mathbf{n}_{\beta\sigma}{\cdot}k_\beta \nabla T_\beta = \mathbf{n}_{\beta\sigma}{\cdot}k_\sigma \nabla T_\sigma \,, \quad \text{on } A_{\beta\sigma} \,, \tag{4.4}$$

respectively, where $\mathbf{n}_{\beta\sigma}$ is the unit vector normal to the solid–liquid interface, oriented toward the phase σ. In general, the complexity of the system microstructure makes it impossible to determine the local temperature field. The alternative is to derive a macroscopic representation (an equivalent continuum model) which represents insofar as possible the geometry and physics on the scale of the pore. To this end, several homogenisation techniques can be put to work. In the present discussion, the local equations will be upscaled by means of the volume averaging technique [3].

4.2 Local Volume Averages

We consider an averaging volume V (see Fig. 4.1) whose characteristic size r_0 must satisfy the scale separation constraints

$$l_\beta, l_\sigma \ll r_0 \ll L\,, \tag{4.5}$$

where l_β and l_σ are the characteristic lengths of the local geometry in each phase and L is the length scale of macroscopic temperature variations (the system scale). If Ψ_β is a physical quantity associated with phase β (for example, temperature, concentration, velocity, etc.), the *superficial volume average* of Ψ_β over the volume V, defined at the center $\mathbf{x}$ of V, is given by

$$\langle \Psi_\beta \rangle|_{\mathbf{x}} = \frac{1}{V}\int_{V_\beta} \Psi_\beta(\mathbf{x}+\mathbf{y}_\beta)\mathrm{d}V\,, \tag{4.6}$$

where V_β is the volume occupied by the phase β in V. In many cases, the *intrinsic volume average* of Ψ_β turns out to be more representative:

$$\langle \Psi_\beta \rangle^\beta|_{\mathbf{x}} = \frac{1}{V_\beta}\int_{V_\beta} \Psi_\beta(\mathbf{x}+\mathbf{y}_\beta)\mathrm{d}V\,. \tag{4.7}$$

To simplify the notation, we shall just write $\langle \Psi_\beta \rangle = \langle \Psi_\beta \rangle|_{\mathbf{x}}$. The intrinsic and superficial averages are related by

$$\langle \Psi_\beta \rangle = \varepsilon_\beta \langle \Psi_\beta \rangle^\beta\,, \tag{4.8}$$

where ε_β is the volume fraction of phase β (corresponding to the porosity in the case of a saturated porous medium). The averaged conservation equations are obtained by applying the following spatial and temporal differentiation theorems [4]:

$$\langle \nabla \Psi_\beta \rangle = \nabla \langle \Psi_\beta \rangle + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\Psi_\beta \mathrm{d}A \tag{4.9}$$

and

$$\left\langle \frac{\partial \Psi_\beta}{\partial t} \right\rangle = \frac{\partial \langle \Psi_\beta \rangle}{\partial t} - \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot\mathbf{w}_{\beta\sigma}\Psi_\beta \mathrm{d}A\,, \tag{4.10}$$

where $\mathbf{w}_{\beta\sigma}$ is the velocity of the interface $A_{\beta\sigma}$.

4.3 Averaged Equations

In the case of a composite made up of two solid phases σ and β, and when the interface $A_{\beta\sigma}$ between them is not moving, application of the volume averaging theorems (4.9) and (4.10) to (4.1) and (4.2) leads to the expressions

$$\varepsilon_\sigma(\rho c_p)_\sigma \frac{\partial \langle T_\sigma \rangle^\sigma}{\partial t} = \nabla\cdot\left[k_\sigma\left(\varepsilon_\sigma \nabla\langle T_\sigma\rangle^\sigma + \langle T_\sigma\rangle^\sigma \nabla\varepsilon_\sigma + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta} T_\sigma \mathrm{d}A\right)\right] + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta}\cdot k_\sigma \nabla T_\sigma \mathrm{d}A \tag{4.11}$$

and

$$\varepsilon_\beta(\rho c_p)_\beta \frac{\partial \langle T_\beta \rangle^\beta}{\partial t} = \nabla\cdot\left[k_\beta\left(\varepsilon_\beta \nabla\langle T_\beta\rangle^\beta + \langle T_\beta\rangle^\beta \nabla\varepsilon_\beta + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma} T_\beta \mathrm{d}A\right)\right] + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot k_\beta \nabla T_\beta \mathrm{d}A\,. \tag{4.12}$$

Note that the integrals in (4.11) and (4.12) depend on the local temperatures T_σ and T_β at the field point $\mathbf{r}$. These can be decomposed in the form [5]

$$T_\sigma = \langle T_\sigma\rangle^\sigma|_{\mathbf{r}} + \widetilde{T}_\sigma\,, \tag{4.13}$$

and

$$T_\beta = \langle T_\beta\rangle^\beta|_{\mathbf{r}} + \widetilde{T}_\beta\,, \tag{4.14}$$

where $\widetilde{T}_\sigma$ and $\widetilde{T}_\beta$ are the local temperature deviations. Introducing these expressions into (4.11) and (4.12) yields

$$\varepsilon_\sigma(\rho c_p)_\sigma \frac{\partial \langle T_\sigma \rangle^\sigma}{\partial t} = \nabla\cdot\left[k_\sigma\left(\varepsilon_\sigma \nabla\langle T_\sigma\rangle^\sigma + \langle T_\sigma\rangle^\sigma \nabla\varepsilon_\sigma + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta} \langle T_\sigma\rangle^\sigma|_{\mathbf{r}} \mathrm{d}A\right)\right] + \nabla\cdot\left(k_\sigma \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta}\widetilde{T}_\sigma \mathrm{d}A\right) + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta}\cdot k_\sigma \nabla T_\sigma \mathrm{d}A \tag{4.15}$$

and

$$\varepsilon_\beta(\rho c_p)_\beta \frac{\partial \langle T_\beta \rangle^\beta}{\partial t} = \nabla\cdot\left[k_\beta\left(\varepsilon_\beta \nabla\langle T_\beta\rangle^\beta + \langle T_\beta\rangle^\beta \nabla\varepsilon_\beta + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma} \langle T_\beta\rangle^\beta|_{\mathbf{r}} \mathrm{d}A\right)\right] + \nabla\cdot\left(k_\beta \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\widetilde{T}_\beta \mathrm{d}A\right) + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot k_\beta \nabla T_\beta \mathrm{d}A\,. \tag{4.16}$$

Furthermore, it can be shown quite generally that, when the following constraints are satisfied [3]

$$l_\gamma \ll r_0\,, \quad \gamma = \beta,\sigma\,, \quad {r_0}^2 \ll L_\varepsilon L_{T1}\,, \tag{4.17}$$

where L_ε and L_{T1} are defined by

$$\nabla\varepsilon_\gamma = O\left(\frac{\Delta\varepsilon_\gamma}{L_\varepsilon}\right), \qquad \nabla\nabla\langle T_\gamma\rangle^\gamma = O\left(\frac{\nabla\langle T_\gamma\rangle^\gamma}{L_{T1}}\right), \quad \gamma=\beta,\sigma, \tag{4.18}$$

we have

$$\langle T_\gamma\rangle^\gamma|_{\mathbf{r}} \sim \langle T_\gamma\rangle^\gamma|_{\mathbf{x}} = \langle T_\gamma\rangle^\gamma. \tag{4.19}$$

Under such conditions, it is relatively easy to show that

$$\frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta}\langle T_\sigma\rangle^\sigma|_{\mathbf{x}}\mathrm{d}A = -\langle T_\sigma\rangle^\sigma\nabla\varepsilon_\sigma \tag{4.20}$$

and

$$\frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\langle T_\beta\rangle^\beta|_{\mathbf{x}}\mathrm{d}A = -\langle T_\beta\rangle^\beta\nabla\varepsilon_\beta. \tag{4.21}$$

Consequently, (4.15) and (4.16) simplify to

$$\begin{aligned}\varepsilon_\sigma(\rho c_p)_\sigma\frac{\partial\langle T_\sigma\rangle^\sigma}{\partial t} &= \nabla\cdot\left[k_\sigma\left(\varepsilon_\sigma\nabla\langle T_\sigma\rangle^\sigma + \frac{1}{V}\int_{A_{\beta\sigma}}\mathbf{n}_{\sigma\beta}\widetilde{T}_\sigma\mathrm{d}A\right)\right]\\ &\quad + \frac{1}{V}\int_{A_{\beta\sigma}}\mathbf{n}_{\sigma\beta}\cdot k_\sigma\nabla T_\sigma\mathrm{d}A\end{aligned} \tag{4.22}$$

and

$$\begin{aligned}\varepsilon_\beta(\rho c_p)_\beta\frac{\partial\langle T_\beta\rangle^\beta}{\partial t} &= \nabla\cdot\left[k_\beta\left(\varepsilon_\beta\nabla\langle T_\beta\rangle^\beta + \frac{1}{V}\int_{A_{\beta\sigma}}\mathbf{n}_{\beta\sigma}\widetilde{T}_\beta\mathrm{d}A\right)\right]\\ &\quad + \frac{1}{V}\int_{A_{\beta\sigma}}\mathbf{n}_{\beta\sigma}\cdot k_\beta\nabla T_\beta\mathrm{d}A.\end{aligned} \tag{4.23}$$

In most conduction problems, the transfer mechanisms can be described using a model with a single energy conservation equation by applying the principle of *local thermal equilibrium* to be discussed in the next section.

4.3.1 Local Thermal Equilibrium and the Single-Equation Model

The characteristic feature of the notion of local thermal equilibrium is expressed by the approximation

$$\langle T_\beta\rangle^\beta = \langle T_\sigma\rangle^\sigma. \tag{4.24}$$

When (4.24) is satisfied, one can add (4.22) and (4.23) to obtain the non-closed macroscopic expression

$$\langle\rho\rangle C_p \frac{\partial\langle T\rangle}{\partial t} = \nabla\cdot\Bigg(\varepsilon_\beta k_\beta \nabla\langle T_\beta\rangle^\beta + \varepsilon_\sigma k_\sigma \nabla\langle T_\sigma\rangle^\sigma + \frac{k_\beta}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\widetilde{T}_\beta \mathrm{d}A + \frac{k_\sigma}{V}\int_{A_{\sigma\beta}} \mathbf{n}_{\sigma\beta}\widetilde{T}_\sigma \mathrm{d}A\Bigg), \tag{4.25}$$

where

$$\langle\rho\rangle C_p = \varepsilon_\beta(\rho c_p)_\beta + \varepsilon_\sigma(\rho c_p)_\sigma \tag{4.26}$$

and

$$\langle T\rangle = \varepsilon_\beta\langle T_\beta\rangle^\beta + \varepsilon_\sigma\langle T_\sigma\rangle^\sigma . \tag{4.27}$$

Under these conditions, (4.24) can be written in the form

$$\langle T\rangle = \langle T_\beta\rangle^\beta = \langle T_\sigma\rangle^\sigma . \tag{4.28}$$

Naturally, the local thermal equilibrium hypothesis depends on a certain number of conditions being satisfied, as examined in detail by Quintard and Whitaker [2, 3]. The aim in the present discussion is not to examine the details of their analysis, but rather to outline the main conclusions. To sum up, the local thermal equilibrium hypothesis is satisfied whenever at least one of the following three conditions holds:

- One of the volume fractions ε_β, ε_σ is zero.
- The phases σ and β have rather similar physical properties.
- The ratio $(l_\beta/L)^2$ tends to zero.

4.3.2 Deviation Equations

The aim here is to present a closed form of (4.25) by determining the deviation fields $\widetilde{T}_\beta$ and $\widetilde{T}_\sigma$. To do this, the first step is to write down equations for the deviations. Since the problems are the same for both phases, we shall only consider $\widetilde{T}_\beta$ in the following. We begin by dividing (4.23) by the volume fraction ε_β to give

$$(\rho c_p)_\beta \frac{\partial\langle T_\beta\rangle^\beta}{\partial t} = \nabla\cdot\left(k_\beta\nabla\langle T_\beta\rangle^\beta\right) + \varepsilon_\beta^{-1}\nabla\varepsilon_\beta\cdot k_\beta\nabla\langle T_\beta\rangle^\beta + \varepsilon_\beta^{-1}\nabla\cdot\left(\frac{k_\beta}{V}\int_{A_{\beta\sigma}}\mathbf{n}_{\beta\sigma}\widetilde{T}_\beta\mathrm{d}A\right) + \frac{\varepsilon_\beta^{-1}}{V}\int_{A_{\beta\sigma}}\mathbf{n}_{\beta\sigma}\cdot k_\beta\nabla T_\beta\mathrm{d}A . \tag{4.29}$$

Equation (4.29) is then subtracted from the local equation (4.2), in which the decomposition (4.14) has been inserted, to obtain

$$(\rho c_p)_\beta \frac{\partial \widetilde{T}_\beta}{\partial t} = \nabla\cdot\left(k_\beta \nabla \widetilde{T}_\beta\right) - \varepsilon_\beta^{-1}\nabla\varepsilon_\beta\cdot k_\beta \nabla\langle T_\beta\rangle^\beta \tag{4.30}$$
$$-\varepsilon_\beta^{-1}\nabla\cdot\left(\frac{k_\beta}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\widetilde{T}_\beta \mathrm{d}A\right) - \frac{\varepsilon_\beta^{-1}}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot k_\beta \nabla T_\beta \mathrm{d}A\,.$$

Given (4.21), the last term of (4.30) corresponding to the interface flux can be written in the form

$$\frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot k_\beta \nabla T_\beta \mathrm{d}A = -\nabla\varepsilon_\beta\cdot k_\beta\nabla\langle T_\beta\rangle^\beta + \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot k_\beta\nabla\widetilde{T}_\beta \mathrm{d}A\,, \tag{4.31}$$

whence (4.30) takes on the simplified form

$$(\rho c_p)_\beta \frac{\partial \widetilde{T}_\beta}{\partial t} = \nabla\cdot\left(k_\beta \nabla \widetilde{T}_\beta\right) - \varepsilon_\beta^{-1}\nabla\cdot\left(\frac{k_\beta}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\widetilde{T}_\beta \mathrm{d}A\right) - \frac{\varepsilon_\beta^{-1}}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot k_\beta \nabla \widetilde{T}_\beta \mathrm{d}A\,. \tag{4.32}$$

The third term in (4.32) is a non-local contribution to the deviation field. Given the following orders of magnitude

$$\varepsilon_\beta^{-1}\nabla\cdot\left(\frac{k_\beta}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\widetilde{T}_\beta \mathrm{d}A\right) = O\left(\frac{\varepsilon_\beta^{-1}k_\beta \widetilde{T}_\beta}{l_\beta L}\right) \tag{4.33}$$

and

$$\nabla\cdot\left(k_\beta\nabla\widetilde{T}_\beta\right) = O\left(\frac{k_\beta\widetilde{T}_\beta}{l_\beta^2}\right)\,, \tag{4.34}$$

and taking into account the scale constraint (4.5), this non-local term can be neglected. Under these conditions, (4.32) reduces to

$$(\rho c_p)_\beta \frac{\partial \widetilde{T}_\beta}{\partial t} = \nabla\cdot\left(k_\beta \nabla \widetilde{T}_\beta\right) - \frac{\varepsilon_\beta^{-1}}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\cdot k_\beta \nabla \widetilde{T}_\beta \mathrm{d}A\,. \tag{4.35}$$

Finally, it should be noted that, although the conduction phenomena considered here are not steady state on the macroscopic scale, the deviation problems can be treated as quasi-steady provided that the following inequality is satisfied:

$$\frac{\alpha_\beta t^*}{l_\beta^2} \gg 1\,, \tag{4.36}$$

where t^* is the characteristic time scale. Under these conditions, the deviation equations for $\widetilde{T}_\beta$ and $\widetilde{T}_\sigma$ become

$$\nabla\cdot\left(k_\sigma\nabla\widetilde{T}_\sigma\right) = \frac{\varepsilon_\sigma^{-1}}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta}\cdot k_\sigma\nabla\widetilde{T}_\sigma \mathrm{d}A \tag{4.37}$$

and

$$\nabla\cdot\left(k_\beta\nabla\widetilde{T}_\beta\right)=\frac{\varepsilon_\beta^{-1}}{V}\int_{A_{\beta\sigma}}\mathbf{n}_{\beta\sigma}\cdot k_\beta\nabla\widetilde{T}_\beta\mathrm{d}A\,, \tag{4.38}$$

with

$$\widetilde{T}_\beta=\widetilde{T}_\sigma+\langle T_\sigma\rangle^\sigma-\langle T_\beta\rangle^\beta\,,\quad\text{on } A_{\beta\sigma}\,, \tag{4.39}$$

and

$$-\mathbf{n}_{\beta\sigma}\cdot k_\beta\nabla\widetilde{T}_\beta=-\mathbf{n}_{\beta\sigma}\cdot k_\sigma\nabla\widetilde{T}_\sigma+\mathbf{n}_{\beta\sigma}\cdot k_\beta\nabla\langle T_\beta\rangle^\beta-\mathbf{n}_{\beta\sigma}\cdot k_\sigma\nabla\langle T_\sigma\rangle^\sigma\,,\quad\text{on } A_{\beta\sigma}\,. \tag{4.40}$$

The system of equations (4.37–4.40) is supplemented by two further boundary conditions:

$$\widetilde{T}_\sigma=g(\mathbf{r},t)\,,\quad\text{on } A_{\sigma\mathrm{e}}\,, \tag{4.41}$$

and

$$\widetilde{T}_\beta=f(\mathbf{r},t)\,,\quad\text{on } A_{\beta\mathrm{e}}\,, \tag{4.42}$$

where surfaces $A_{\sigma\mathrm{e}}$ and $A_{\beta\mathrm{e}}$ are the boundaries of the volume averaging region V in contact with the environment. At this stage, the functions f and g are unknown.

We observe that the boundary conditions (4.39) and (4.40) contain three sources of deviations, viz., $\langle T_\sigma\rangle^\sigma-\langle T_\beta\rangle^\beta$, $\nabla\langle T_\beta\rangle^\beta$, and $\nabla\langle T_\sigma\rangle^\sigma$. Assuming local thermal equilibrium, these three terms can be replaced by the single source term $\nabla\langle T\rangle$. Under these conditions, (4.39) and (4.40) can be written in the form

$$\widetilde{T}_\beta=\widetilde{T}_\sigma\,,\quad\text{on } A_{\beta\sigma}\,, \tag{4.43}$$

and

$$-\mathbf{n}_{\beta\sigma}\cdot k_\beta\nabla\widetilde{T}_\beta=-\mathbf{n}_{\beta\sigma}\cdot k_\sigma\nabla\widetilde{T}_\sigma+\mathbf{n}_{\beta\sigma}\cdot(k_\beta-k_\sigma)\nabla\langle T\rangle\,,\quad\text{on } A_{\beta\sigma}\,. \tag{4.44}$$

Generally speaking, the complexity of the local geometry (microstructure) requires one to choose a *periodic* volume V that is representative of the composite system. In this case, the boundary conditions at the surfaces $A_{\sigma\mathrm{e}}$ and $A_{\beta\mathrm{e}}$ take the form

$$\widetilde{T}_\sigma(\mathbf{r}+l_i)=\widetilde{T}_\sigma(\mathbf{r})\,,\quad\text{on } A_{\sigma\mathrm{e}}\,, \tag{4.45}$$

and

$$\widetilde{T}_\beta(\mathbf{r}+l_i)=\widetilde{T}_\beta(\mathbf{r})\,,\quad\text{on } A_{\beta\mathrm{e}}\,, \tag{4.46}$$

where l_i, $i=1,2,3$ are the three vectors defining the unit cell. Finally, it can be shown that the average deviation is zero [3], whence

$$\langle\widetilde{T}_\sigma\rangle^\sigma=0\,,\qquad\langle\widetilde{T}_\beta\rangle^\beta=0\,. \tag{4.47}$$

The second step toward setting up a closed form of (4.25) is to transform the deviation problem into a closure problem.

4.3.3 Closure Problem

In the framework of local thermal equilibrium, we may assume to a first approximation that $\widetilde{T}_\sigma$ and $\widetilde{T}_\beta$ are proportional to the source term $\nabla\langle T\rangle$, i.e.,

$$\widetilde{T}_\sigma = \mathbf{b}_\sigma \cdot \nabla\langle T\rangle \tag{4.48}$$

and

$$\widetilde{T}_\beta = \mathbf{b}_\beta \cdot \nabla\langle T\rangle \,, \tag{4.49}$$

where $\mathbf{b}_\sigma$ and $\mathbf{b}_\beta$ are the closure variables. Introducing (4.48) and (4.49) into the deviation equations (4.37) and (4.38), the system (4.43–4.46) leads to the closure problem

$$k_\sigma \nabla^2 \mathbf{b}_\sigma = \frac{\varepsilon_\sigma^{-1}}{V} \int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta} \cdot k_\sigma \nabla \mathbf{b}_\sigma \mathrm{d}A \tag{4.50}$$

and

$$k_\beta \nabla^2 \mathbf{b}_\beta = \frac{\varepsilon_\beta^{-1}}{V} \int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma} \cdot k_\beta \nabla \mathbf{b}_\beta \mathrm{d}A \,, \tag{4.51}$$

with

$$\mathbf{b}_\beta = \mathbf{b}_\sigma \,, \quad \text{on } A_{\beta\sigma} \,, \tag{4.52}$$

$$-\mathbf{n}_{\beta\sigma} \cdot k_\beta \nabla \mathbf{b}_\beta = -\mathbf{n}_{\beta\sigma} \cdot k_\sigma \nabla \mathbf{b}_\sigma + \mathbf{n}_{\beta\sigma}(k_\beta - k_\sigma) \,, \quad \text{on } A_{\beta\sigma} \,, \tag{4.53}$$

$$\mathbf{b}_\sigma(\mathbf{r} + l_i) = \mathbf{b}_\sigma(\mathbf{r}) \,, \quad \text{on } A_{\sigma\mathrm{e}} \,, \tag{4.54}$$

and

$$\mathbf{b}_\beta(\mathbf{r} + l_i) = \mathbf{b}_\beta(\mathbf{r}) \,, \quad \text{on } A_{\beta\mathrm{e}} \,. \tag{4.55}$$

Furthermore, given (4.47), we have

$$\langle \mathbf{b}_\sigma \rangle^\sigma = 0 \,, \qquad \langle \mathbf{b}_\beta \rangle^\beta = 0 \,. \tag{4.56}$$

For present purposes, we shall not tackle the general case of solving (4.50–4.56) for an arbitrary unit cell. We shall instead focus on the simpler case of a *symmetric* unit cell. For this case, using the symmetry conditions and the divergence theorem, we obtain

$$\frac{1}{V} \int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta} \cdot k_\sigma \nabla \mathbf{b}_\sigma \mathrm{d}A = \frac{1}{V} \int_{V_\sigma} k_\sigma \nabla^2 \mathbf{b}_\sigma \mathrm{d}A = 0 \tag{4.57}$$

and

$$\frac{1}{V} \int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma} \cdot k_\beta \nabla \mathbf{b}_\beta \mathrm{d}A = \frac{1}{V} \int_{V_\beta} k_\beta \nabla^2 \mathbf{b}_\beta \mathrm{d}A = 0 \,. \tag{4.58}$$

The closure problem (4.50–4.56) then assumes the simplified form [6]

$$\nabla^2 \mathbf{b}_\sigma = 0 \tag{4.59}$$

and

$$\nabla^2 \mathbf{b}_\beta = 0 \,, \tag{4.60}$$

with

$$\mathbf{b}_\beta = \mathbf{b}_\sigma \,, \quad \text{on } A_{\beta\sigma} \,, \tag{4.61}$$

$$-\mathbf{n}_{\beta\sigma}\cdot\nabla\mathbf{b}_\beta = -\mathbf{n}_{\beta\sigma}\cdot\kappa\nabla\mathbf{b}_\sigma + \mathbf{n}_{\beta\sigma}(1-\kappa) \,, \quad \text{on } A_{\beta\sigma} \,, \tag{4.62}$$

$$\mathbf{b}_\gamma(\mathbf{r}+l_i) = \mathbf{b}_\gamma(\mathbf{r}) \,, \quad \text{on } A_{\gamma\mathrm{e}} \,, \quad \gamma = \beta, \sigma \,, \tag{4.63}$$

and

$$\langle \mathbf{b}_\sigma \rangle^\sigma = 0 \,, \qquad \langle \mathbf{b}_\beta \rangle^\beta = 0 \,, \tag{4.64}$$

where κ is the ratio of the conductivities in the two phases, viz.,

$$\kappa = k_\sigma / k_\beta \,. \tag{4.65}$$

When $\kappa = 0$, the problem is precisely as would be obtained for a diffusion problem in a porous medium with zero diffusion in the solid phase (phase σ) [3].

4.3.4 Closed Form

The non-closed form of the single-equation model is

$$\langle\rho\rangle C_p \frac{\partial\langle T\rangle}{\partial t} = \nabla\cdot\Bigg(\varepsilon_\beta k_\beta \nabla\langle T_\beta\rangle^\beta + \varepsilon_\sigma k_\sigma \nabla\langle T_\sigma\rangle^\sigma + \frac{k_\beta}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\widetilde{T}_\beta \mathrm{d}A + \frac{k_\sigma}{V}\int_{A_{\sigma\beta}} \mathbf{n}_{\sigma\beta}\widetilde{T}_\sigma \mathrm{d}A\Bigg) \,. \tag{4.66}$$

Substituting (4.48) and (4.49) into (4.66), with

$$\frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\sigma\beta}\mathrm{d}A = 0 \,, \qquad \frac{1}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\mathrm{d}A = 0 \,, \tag{4.67}$$

leads to the following closed macroscopic form:

$$\langle\rho\rangle C_p \frac{\partial\langle T\rangle}{\partial t} = \nabla\cdot\big(\mathsf{K}_{\mathrm{eff}}\cdot\nabla\langle T\rangle\big) \,, \tag{4.68}$$

where

$$\mathsf{K}_{\mathrm{eff}} = \big(\varepsilon_\beta k_\beta + \varepsilon_\sigma k_\sigma\big) + \frac{k_\beta - k_\sigma}{V}\int_{A_{\beta\sigma}} \mathbf{n}_{\beta\sigma}\mathbf{b}_\beta \mathrm{d}A \,. \tag{4.69}$$

It should be noted that that the last term in (4.69) corresponds to a contribution related to the tortuosity of the interface $A_{\beta\sigma}$. Determination of the effective conductivity tensor (4.69) for a given composite structure thus depends on the field of the closure variable $\mathbf{b}_\beta$ which solves the differential system (4.59–4.64) for the same structure. Finally, it can be shown that the effective conductivity tensor (4.69) is

symmetric, i.e.,

$$\mathsf{K}_{\text{eff}} = \mathsf{K}_{\text{eff}}^{\text{T}} , \tag{4.70}$$

where the superscript T denotes transposition.

4.3.5 Local Thermal Non-Equilibrium

When the local thermal equilibrium hypothesis is not satisfied, Quintard and Whitaker [2] propose a closed form of (4.22) and (4.23) which leads to the following two-equation model:

$$\varepsilon_\sigma (\rho c_p)_\sigma \frac{\partial \langle T_\sigma \rangle^\sigma}{\partial t} = \nabla \cdot \left(\mathbf{K}_{\sigma\beta} \cdot \nabla \langle T_\beta \rangle^\beta + \mathsf{K}_{\sigma\sigma} \cdot \nabla \langle T_\sigma \rangle^\sigma \right) - a_v h \left(\langle T_\sigma \rangle^\sigma - \langle T_\beta \rangle^\beta \right) \tag{4.71}$$

and

$$\varepsilon_\beta (\rho c_p)_\beta \frac{\partial \langle T_\beta \rangle^\beta}{\partial t} = \nabla \cdot \left(\mathsf{K}_{\beta\beta} \cdot \nabla \langle T_\beta \rangle^\beta + \mathsf{K}_{\beta\sigma} \cdot \nabla \langle T_\sigma \rangle^\sigma \right) - a_v h \left(\langle T_\beta \rangle^\beta - \langle T_\sigma \rangle^\sigma \right) . \tag{4.72}$$

It can be shown in general that

$$\mathsf{K}_{\beta\sigma} = \mathsf{K}_{\sigma\beta} \tag{4.73}$$

and also

$$\mathsf{K}_{\beta\sigma} \ll \mathsf{K}_{\beta\beta} , \quad \mathsf{K}_{\sigma\sigma} . \tag{4.74}$$

Under these conditions, the two-equation system becomes

$$\varepsilon_\sigma (\rho c_p)_\sigma \frac{\partial \langle T_\sigma \rangle^\sigma}{\partial t} = \nabla \cdot (\mathsf{K}_{\sigma\sigma} \cdot \nabla \langle T_\sigma \rangle^\sigma) - a_v h \left(\langle T_\sigma \rangle^\sigma - \langle T_\beta \rangle^\beta \right) \tag{4.75}$$

and

$$\varepsilon_\beta (\rho c_p)_\beta \frac{\partial \langle T_\beta \rangle^\beta}{\partial t} = \nabla \cdot \left(\mathsf{K}_{\beta\beta} \cdot \nabla \langle T_\beta \rangle^\beta \right) - a_v h \left(\langle T_\beta \rangle^\beta - \langle T_\sigma \rangle^\sigma \right) , \tag{4.76}$$

where h is the effective interface exchange constant, obtained by solving an associated closure problem.

References

1. R.G. Carbonell, S. Whitaker: *Heat and Mass Transfer in Porous Media* (Martinus Nijkoff, Dordrecht, 1984) pp. 121–198
2. M. Quintard, S. Whitaker: Adv. Heat Transfer **23**, 369 (1993)
3. S. Whitaker: *The Method of Volume Averaging*, Vol. 13 (Kluwer Academic Publishers, 1999)
4. S. Whitaker: Ind. Eng. Chem **12**, 12 (1969)
5. W.G. Gray: Chem. Eng. Sci. **3**, 229 (1975)
6. I. Nozad, R.G. Carbonell, S. Whitaker: Chem. Eng. Sci. **40**, 843 (1985)

Chapter 5
Heat Conduction in Composites

Jean-Yves Duquesne

In this chapter we present the basics of heat transfer in composites. The first approach, called the effective medium method, averages in a suitable way over the conductivities of the constituents. This gives good results for sufficiently large particles. However, when the particles are smaller than the mean free path of the excitations carrying the heat, the implicit assumptions of the effective medium theory are no longer justified. One must then look at the way the particles scatter the energy carriers.

5.1 Microcomposites and Effective Media

5.1.1 Taking Averages

Consider a medium comprising a matrix with particle inclusions. Figure 5.1 shows the situation schematically. It is assumed here that heat conduction in the matrix and particles can be described by Fourier's law, with the same conductivities as for the corresponding bulk materials:

$$\mathbf{q} = -\Lambda \nabla T \,, \tag{5.1}$$

where $\mathbf{q}$ is the heat flux, Λ the thermal conductivity, and T the temperature.

This assumes implicitly that the mean free paths of the energy carrying excitations are much shorter than the characteristic dimensions of the matrix and particles. More precisely, if l_{m} and l_{p} are the mean free paths in the bulk materials making up the matrix and particles, the distances between particles must be large compared with l_{m}, and the particle sizes must be large compared with l_{p}.

The presentation here is based on the accounts in [1, 2]. The aim is to relate the incoming heat flux (equal to the outgoing heat flux in the steady-state regime) to the temperature gradient which can be measured from outside the sample, viz., $(T_2 - T_1)/L$, via an effective conductivity Λ^*. Here L is the length of the sample. Let

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_5, © Springer-Verlag Berlin Heidelberg 2009

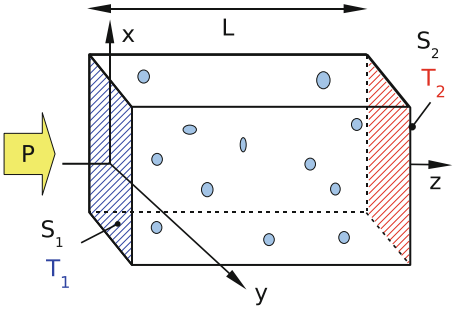

Fig. 5.1 Sample composite under a heat gradient, where T_1 and T_2 are the temperatures imposed on faces S_1 and S_2, and P is the incoming power

S be the area of the faces S_1 and S_2, and z the axis perpendicular to those faces. Then P and P/S are the power and flux coming in through the face S_1. Further, V, V_m, and V_p denote the volume of the whole sample, the volume of the matrix, and the total volume of the particles, respectively, and Λ_m, Λ_p are the thermal conductivities of the matrix and particles, respectively. The volume fractions η_m and η_p occupied by matrix and particles, respectively, are clearly

$$\eta_i = \frac{V_i}{V}\,, \qquad i = \mathrm{m}, \mathrm{p}\,. \tag{5.2}$$

We also define the average value of any quantity X over the whole sample by

$$\langle X \rangle = \frac{1}{V} \iiint_V X \mathrm{d}v\,, \tag{5.3}$$

and the average value of the quantity X over the matrix or the particles by

$$\langle X \rangle_i = \frac{1}{V_i} \iiint_{V_i} X \mathrm{d}v\,, \qquad i = \mathrm{m}, \mathrm{p}\,. \tag{5.4}$$

Temperatures T_1 and T_2 are imposed on the end faces S_1 and S_2 of the sample. A heat flux $\mathbf{q}$ is then set up. It is assumed that there is no flux through the side walls. Then, by the definition of Λ^*, we have

$$\frac{P}{S} = -\Lambda^* \frac{T_2 - T_1}{L}\,, \tag{5.5}$$

with

$$P = -\iint_{S_1} \mathbf{q}\cdot\mathrm{d}\mathbf{S} = \iint_{S_2} \mathbf{q}\cdot\mathrm{d}\mathbf{S}\,. \tag{5.6}$$

Let us relate the incoming flux to the volume average of q_z. We have (see Appendix A)

$$\iiint_V q_z \mathrm{d}v = \iint_{\partial V} z\mathbf{q}\cdot\mathrm{d}\mathbf{S}\,. \tag{5.7}$$

Note that, at the side faces, $\mathbf{q}$ is parallel to the surface. The above surface integral thus reduces to the integral over the end faces S_1 and S_2. Furthermore, the integral over S_1 is zero because $z = 0$ there, whence

$$\iint_{\partial V} z\mathbf{q}\cdot\mathrm{d}\mathbf{S} = L\iint_{S_2} \mathbf{q}\cdot\mathrm{d}\mathbf{S}\,, \tag{5.8}$$

and

$$\iiint_V q_z \mathrm{d}v = L\iint_{S_2} \mathbf{q}\cdot\mathrm{d}\mathbf{S} = LP\,. \tag{5.9}$$

The relation we seek between $\langle q_z\rangle$ and the incoming flux is therefore

$$\langle q_z\rangle = \frac{1}{V}\iiint_V q_z \mathrm{d}v = \frac{P}{S}\,. \tag{5.10}$$

Consequently,

$$\langle q_z\rangle = -\Lambda^* \frac{T_2 - T_1}{L}\,. \tag{5.11}$$

The average value of q_z can also be obtained by decomposing the average over the volumes occupied by the matrix and particles:

$$\langle q_z\rangle = \eta_\mathrm{m}\langle q_z\rangle_\mathrm{m} + \eta_\mathrm{p}\langle q_z\rangle_\mathrm{p}\,. \tag{5.12}$$

Applied locally in the matrix ($i = \mathrm{m}$) and in the particles ($i = \mathrm{p}$), the Fourier law gives

$$q_z = -\Lambda_i \frac{\mathrm{d}T}{\mathrm{d}z}\,. \tag{5.13}$$

Equations (5.12) and (5.13) lead to

$$\langle q_z\rangle = -\eta_\mathrm{m}\Lambda_\mathrm{m}\left\langle \frac{\mathrm{d}T}{\mathrm{d}z}\right\rangle_\mathrm{m} - \eta_\mathrm{p}\Lambda_\mathrm{p}\left\langle \frac{\mathrm{d}T}{\mathrm{d}z}\right\rangle_\mathrm{p}\,. \tag{5.14}$$

It can also be shown (see Appendix B) that

$$\frac{T_2 - T_1}{L} = \eta_\mathrm{m}\left\langle \frac{\mathrm{d}T}{\mathrm{d}z}\right\rangle_\mathrm{m} + \eta_\mathrm{p}\left\langle \frac{\mathrm{d}T}{\mathrm{d}z}\right\rangle_\mathrm{p} - \eta_\mathrm{p}J_z\,, \tag{5.15}$$

where J_z takes into account temperature discontinuities at the particle–matrix interfaces, due to the interface resistance. In fact,

$$J_z = \frac{1}{V_\mathrm{p}}\iint_I (T_\mathrm{p} - T_\mathrm{m}) n_z \mathrm{d}s\,, \tag{5.16}$$

where I represents the totality of particle–matrix interface and $\mathbf{n}$ is a unit normal vector to I, oriented from the particle out into the matrix. From (5.14) and (5.15), we may now write

$$\langle q_z \rangle = -\Lambda_\mathrm{m} \frac{T_2 - T_1}{L} - \eta_\mathrm{p}(\Lambda_\mathrm{p} - \Lambda_\mathrm{m}) \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_\mathrm{p} - \eta_\mathrm{p} \Lambda_\mathrm{m} J_z \,. \tag{5.17}$$

Using (5.11), we now have

$$\Lambda^* \frac{T_2 - T_1}{L} = \Lambda_\mathrm{m} \frac{T_2 - T_1}{L} + \eta_\mathrm{p}(\Lambda_\mathrm{p} - \Lambda_\mathrm{m}) \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_\mathrm{p} + \eta_\mathrm{p} \Lambda_\mathrm{m} J_z \,. \tag{5.18}$$

5.1.2 Particle Shell

The unknowns remaining in (5.18) are $\langle \mathrm{d}T/\mathrm{d}z \rangle_\mathrm{p}$ and J_z. In order to evaluate these, the particles are assumed to be spherical, with the same radius a. Furthermore, it is assumed that averages calculated over all the particles are equal to the averages calculated over a single particle immersed in a medium representative of the global properties of the composite. This representative medium is constructed as follows:

- In the immediate vicinity of a particle, there are no other particles, so the representative medium must have the conductivity Λ_m of the matrix.
- At large distances from the particle, the heterogeneous medium is replaced by a homogeneous medium which has precisely the effective conductivity Λ^* we are trying to calculate.

Figure 5.2 illustrates what is happening. The immediate neighbourhood of a particle is bounded by a sphere with a radius b that depends on the concentration of particles in the matrix. Quite arbitrarily, we require

$$\left(\frac{a}{b}\right)^3 = \eta_\mathrm{p} \,. \tag{5.19}$$

Then b defined in this way is of the order of $d/2$, where d is the average distance between neighbouring particles. As an example, $b \approx 0.62d$ and $b \approx 0.55d$ for cubic and hexagonal particle lattices, respectively. In the model discussed here, the particle is considered to be surrounded by a shell of conductivity Λ_m, while the whole thing is immersed in the effective medium whose properties (in particular the conductivity Λ^*) we hope to establish.

Now consider the system comprising a single particle with shell immersed in the effective medium (see Fig. 5.2 right). The following requirements are made:

- The system must satisfy the same boundary conditions as the composite, namely, temperatures T_1 and T_2 imposed on two parallel planes far from the particle, a distance L apart, perpendicular to the z axis.
- The flux normal to the interfaces is continuous.
- The thermal conductance at the interface between the shell and the effective medium is infinite.
- The thermal conductance at the particle–shell interface is G, defined by

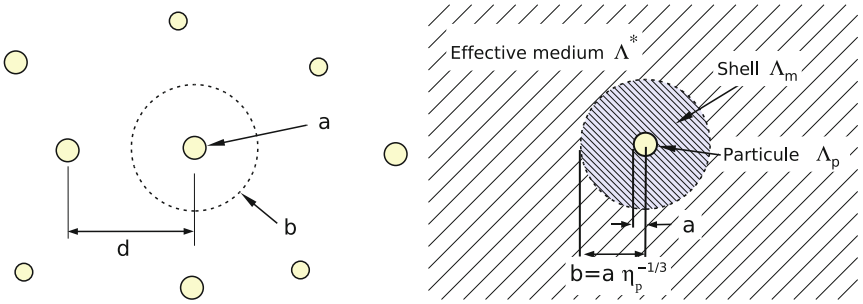

Fig. 5.2 *Left*: Schematic view of a composite. The region bounded by the *dashed circle* defines the immediate vicinity, or shell, of a particle. *Right*: Modelling the composite. a is the particle radius, b the radius of the surrounding shell, d the average distance between neighbouring particles, η_p the volume fraction of particles, and Λ the conductivity

$$q_\mathrm{pm} = -G(T_\mathrm{I,m} - T_\mathrm{I,p}) \,, \tag{5.20}$$

where q_pm is the energy flux normal to the particle–shell interface, and $T_\mathrm{I,m}$ and $T_\mathrm{I,p}$ are the temperatures at the particle–shell interface, in the shell and in the particle, respectively.

The temperature field in the system satisfies Laplace's equation

$$\Delta T = 0 \,. \tag{5.21}$$

The solution has the form

$$T = \begin{cases} Ar\cos\theta & \text{(in the sphere)} \,, \\ (Cr + Dr^{-2})\cos\theta & \text{(in the shell)} \,, \\ (Er + Fr^{-2})\cos\theta & \text{(in the matrix)} \,, \end{cases} \tag{5.22}$$

where r and θ are polar coordinates at the field point. Note that the polar axis is taken parallel to ∇T and the origin at the particle center. It is then a straightforward matter to calculate $\langle \mathrm{d}T/\mathrm{d}z \rangle_\mathrm{p}$ and J_z:

$$\left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_\mathrm{p} = A \,, \tag{5.23}$$

$$J_z = -\frac{\Lambda_\mathrm{p}}{aG} A \,, \tag{5.24}$$

where

$$A = \frac{9\Lambda_m \Lambda^* \dfrac{T_2 - T_1}{L}}{\left[\Lambda_p + 2\Lambda_m\left(1 + \dfrac{\Lambda_p}{aG}\right)\right](\Lambda_m + 2\Lambda^*) + 2\eta_p\left[\Lambda_p - \Lambda_m\left(1 + \dfrac{\Lambda_p}{aG}\right)\right](\Lambda_m - \Lambda^*)} . \tag{5.25}$$

Equations (5.18), (5.23), and (5.24) give a quadratic equation in Λ^* with just one positive root [1]:

$$\Lambda^* = \Lambda_m \frac{\Lambda_p(1+2\alpha) + 2\Lambda_m + 2\eta_p\left[\Lambda_p(1-\alpha) - \Lambda_m\right]}{\Lambda_p(1+2\alpha) + 2\Lambda_m - \eta_p\left[\Lambda_p(1-\alpha) - \Lambda_m\right]} , \tag{5.26}$$

where

$$\alpha = \frac{\Lambda_m}{aG} . \tag{5.27}$$

Then by (5.26), the coefficients specifying the thermal field [see (5.22)] are

$$A = \frac{3\Lambda_m}{\eta_p\left[\Lambda_m + (\alpha-1)\Lambda_p\right] + 2\Lambda_m + (2\alpha+1)\Lambda_p} E , \tag{5.28}$$

$$C = \frac{2\Lambda_m + (2\alpha+1)\Lambda_p}{\eta_p\left[\Lambda_m + (\alpha-1)\Lambda_p\right] + 2\Lambda_m + (2\alpha+1)\Lambda_p} E , \tag{5.29}$$

$$D = \frac{\Lambda_m + (\alpha-1)\Lambda_p}{\eta_p\left[\Lambda_m + (\alpha-1)\Lambda_p\right] + 2\Lambda_m + (2\alpha+1)\Lambda_p} a^3 E , \tag{5.30}$$

$$E = \frac{T_2 - T_1}{L} , \tag{5.31}$$

$$F = 0 . \tag{5.32}$$

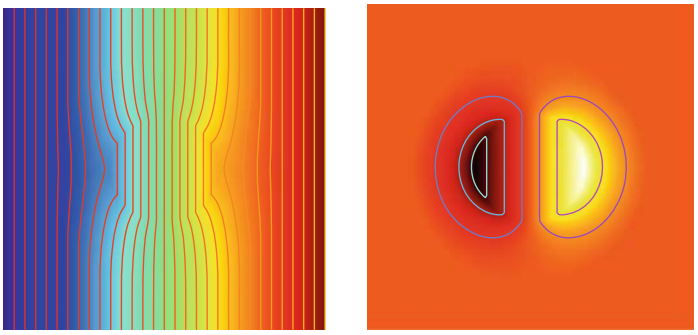

Fig. 5.3 Single particle with shell immersed in a uniform temperature gradient. *Left*: Temperature field and isotherms. *Right*: Temperature field and isotherms after subtracting the imposed field. Calculation parameters: $a = 1$ μm, $b = 2.15$ μm, $\eta_p = 10\%$, $\Lambda_p = 10$ W m^{-1}K^{-1}, $\Lambda_m = 100$ W m^{-1}K^{-1}, $\Lambda^* = 88$ W m^{-1}K^{-1}, G infinite. Image size 6 μm × 6 μm

Figure 5.3 shows the temperature field. As the coefficient F is zero, the range over which the single particle and shell perturbs the imposed field has radius exactly equal to $b = a/\eta_p^{1/3}$. Furthermore, the distance between particle centers is $2b$. This means that the particles in their enclosing shells interact very little with one another. The underlying assumption of the effective medium model, according to which the averages $\langle dT/dz\rangle_p$ and J_z over all the particles are equal to the values calculated for a single particle and shell, is thus fairly well satisfied to a first approximation.

In the limit $\eta_p \to 0$ of small particle concentrations,

$$\Lambda^* = \Lambda_m \left[1 + 3\eta_p \frac{\Lambda_p(1-\alpha) - \Lambda_m}{\Lambda_p(1+2\alpha) + 2\Lambda_m}\right] . \tag{5.33}$$

Depending on the sign of the coefficient of η_p, the conductivity increases or decreases with the particle concentration. If $\Lambda_m < \Lambda_p$, there is a critical particle radius a_c. For $a > a_c$, the conductivity of the composite increases when the concentration increases, while for $a < a_c$, the interface resistance begins to dominate and the conductivity decreases as the concentration increases. We have

$$a_c^{-1} = G(\Lambda_m^{-1} - \Lambda_p^{-1}) . \tag{5.34}$$

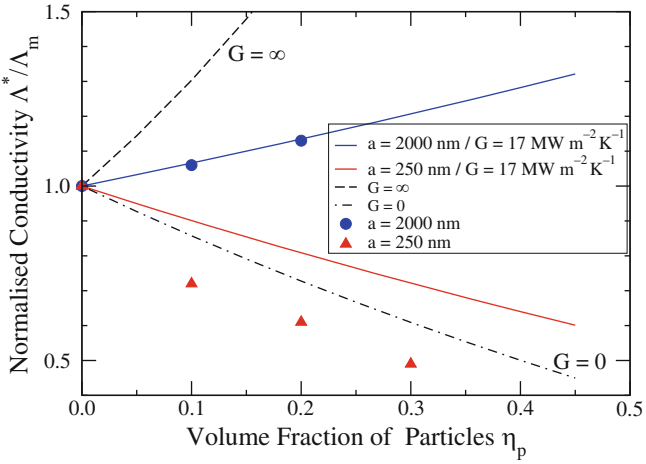

Fig. 5.4 Normalised conductivity Λ^*/Λ_m as a function of the volume fraction of particles. *Circles* and *triangles* represent experimental data for diamond particles in ZnS [3]. *Lines* indicate applications of (5.26). a is the particle radius

5.1.3 Experimental Examples

Figure 5.4 shows experimental results obtained for diamond particles in a ZnS matrix, for two particle sizes, viz., 2 and 0.25 μm [3]. For the large particles, (5.26) can account for the results if an interface conductance $G = 1.7 \times 10^7$ W m^{-2}K^{-1} is introduced. (Note that ZnS and diamond have conductivities 17.4 W m^{-1}K^{-1} and 600 W m^{-1}K^{-1}, respectively.) This value implies a critical radius $a_c = 1$ μm. The conduction is indeed observed to decrease for the smaller particles when their concentration increases. However, (5.26) is unable to approximate the experimental results for the smaller particles, even assuming a very high interface resistance ($G = 0$ W m^{-2}K^{-1}). It has been suggested that this disagreement is due to the non-spherical shape of the small particles [2, 3].

A certain number of interface conductances have been evaluated by interpreting the experimental results in the framework of an effective medium theory: Al/SiC $G = 7 \times 10^8$ W m^{-2}K^{-1} [4], PMMA/Al_2O_3 $G = 3 \times 10^7$ W m^{-2}K^{-1} [5].

5.2 Nanocomposites and Phonon Scattering

5.2.1 Limitations of Effective Medium Theories

Effective medium models make no assumptions about the kind of excitation carrying the heat energy. They assume that heat is transported in the same way in matrix and particles as in the corresponding bulk materials. They thus apply the Fourier law in each medium, with the same conductivities as for the bulk materials. Interfaces are sometimes taken into account and modelled where necessary.

This approach is problematic when one needs to describe composites containing small particles. Indeed, the notions of thermal conductivity, temperature, and so on, are statistical in nature, and defined only for relatively large volumes in which the interaction processes between excitations allow quasi-equilibrium situations to arise. In a homogeneous material, the characteristic size of these volumes is a few times the mean free path of the dominant excitations. At room temperature, the mean free path of the dominant phonons is typically 10–100 nm. At low temperatures, it may reach the centimeter range. So at room temperature, a particle a few nanometers across can be considered small, but at 1 K, a particle a few hundred μm across can be considered small.

Since small particles cannot be handled using the Fourier law with the conductivity of the infinite medium, the effective medium theories discussed in the last section cannot in principle be adapted to the study of nanocomposites. Other methods must be devised. Atomic impurities constitute an extreme case of small particles. Clearly, the thermal conductance of an atom cannot be described using the Fourier law. It is known that the inclusion of impurities within a matrix leads to a reduction in thermal

conductivity. This well known mechanism is due to scattering of phonons by these impurities [6].

The generalisation of this mechanism to larger atomic clusters was proposed long ago [7, 8], and has recently been taken up again [9–12]. We shall discuss this here, restricting to the case where heat is carried by lattice vibrations (phonons).

5.2.2 Kinetic Theory of Heat Transport in Solids

Energy transport by phonons can be treated using the Boltzmann formalism [6]. When there is a temperature gradient, the phonon population is locally out of equilibrium. However, in the steady-state regime, collision processes (phonon–phonon, phonon–defect, etc.) allow the population to remain stable over time. To model these collisions, it is assumed that they induce a return to equilibrium of the perturbed population via some relaxation process. In addition, it is assumed that the characteristic time required for a population of phonons with given wave vector $\mathbf{q}$ to return to equilibrium depends only on how far that particular population is from equilibrium, and not on how far the other phonon populations may be from equilibrium.

Let us give a brief description of the various steps in the calculation [6]. In a homogeneous crystalline material, the phonons can be labelled by the polarisation branch j to which they happen to belong and by their wave vector $\mathbf{q}$. Let ω and $\mathbf{v}$ be their angular frequency and group velocity, respectively. Further, let v and v_z be the modulus and z component of $\mathbf{v}$, and V the sample volume.

The heat flux $\mathbf{Q}$ at a point is

$$\mathbf{Q} = \frac{1}{V} \sum_{\mathbf{q},j} n(\mathbf{q}, j) \hbar \omega(\mathbf{q}, j) \mathbf{v}(\mathbf{q}, j) \,, \tag{5.35}$$

where $n(\mathbf{q}, j)$ is the variation of the phonon distribution function relative to the equilibrium situation. It can be shown that

$$n(\mathbf{q}, j) = -\left[\mathbf{v}(\mathbf{q}, j) \cdot \nabla T\right] \tau(\mathbf{q}, j) \frac{\mathrm{d}N_0}{\mathrm{d}T} \,, \tag{5.36}$$

where T is the temperature, τ is the phonon relaxation time, and N_0 is the equilibrium phonon distribution function, i.e., the Bose–Einstein distribution, given by

$$N_0(\mathbf{q}, j) = \left[\exp \frac{\hbar \omega(\mathbf{q}, j)}{kT} - 1\right]^{-1} \,. \tag{5.37}$$

Assuming that the temperature gradient lies in the z direction, we have

$$\mathbf{Q} = -\frac{\mathrm{d}T}{\mathrm{d}z} \frac{1}{V} \sum_{\mathbf{q},j} \tau(\mathbf{q}, j) \hbar \omega(\mathbf{q}, j) \frac{\mathrm{d}N_0}{\mathrm{d}T} v_z(\mathbf{q}, j) \mathbf{v}(\mathbf{q}, j) \,, \tag{5.38}$$

which yields the following thermal conductivity in the z direction:

$$\Lambda = \frac{1}{V}\sum_{\mathbf{q},j} v_z^2(\mathbf{q},j)\tau(\mathbf{q},j)\hbar\omega(\mathbf{q},j)\frac{\mathrm{d}N_0}{\mathrm{d}T} \,. \tag{5.39}$$

For an isotropic material, the $\mathbf{q}$ dependence only occurs through the wave number q. Replacing the sum over all modes $(\mathbf{q}, j)$ by an integral, and thereby introducing the density of states, (5.39) becomes [6]

$$\Lambda = \frac{1}{3}\sum_j \int_0^{q_{\max}} \mathrm{d}q\, v^2(q,j)\tau(q,j)S(q,j) \,, \tag{5.40}$$

where $S(q,j)\mathrm{d}q$ is the contribution to the specific heat by phonons in the branch j with wave number between q and $q+\mathrm{d}q$. Let $D(q,j)$ be the density of phonons in branch j. Then

$$S(q,j) = k\frac{D(q,j)}{V}\frac{x^2\mathrm{e}^x}{(\mathrm{e}^x-1)^2} \,, \qquad x = \frac{\hbar\omega(q,j)}{kT} \,. \tag{5.41}$$

Consider now a Debye representation of the phonon spectrum:

$$\Lambda = \frac{k^3T^2}{2\pi^2 v_\mathrm{D}\hbar^2}\int_0^{\omega_\mathrm{D}} \tau(\omega,T)g(\omega,T)\mathrm{d}\omega \,, \tag{5.42}$$

with

$$g(\omega,T) = \left(\frac{\hbar\omega}{kT}\right)^4 \frac{\exp(\hbar\omega/kT)}{\left[\exp(\hbar\omega/kT)-1\right]^2} \,, \tag{5.43}$$

where ω_D and v_D are the Debye angular frequency and speed. The spectrum of phonons playing a significant role in heat conduction is determined by the function $\tau(\omega,T)g(\omega,T)$. Typically, these phonons satisfy $\hbar\omega/kT \sim 1$. The relaxation time τ depends on all the various relaxation mechanisms of the phonons. If we assume that these mechanisms consist of those existing in the matrix in the absence of any particles (characterised by a time τ_0), together with the mechanism due to scattering from the particles (characterised by a time τ_P), and if we also assume that these processes are independent, then

$$\tau^{-1}(\omega,T) = \tau_0^{-1}(\omega,T) + \tau_\mathrm{P}^{-1}(\omega,T) \,. \tag{5.44}$$

The time τ_0 itself depends on scattering mechanisms: scattering by the boundaries (τ_B), by Umklapp processes (τ_U), by impurities (τ_I), and so on, whence

$$\tau_0^{-1} = \tau_\mathrm{B}^{-1} + \tau_\mathrm{U}^{-1} + \tau_\mathrm{I}^{-1} + \cdots \,. \tag{5.45}$$

Scattering from the boundaries, impurities, and particles are independent of the temperature. Only the Umklapp processes (τ_U) depend on temperature. We have

$$\tau_{\mathrm{B}}^{-1} \propto l^{-1} \,, \tag{5.46}$$

$$\tau_{\mathrm{I}}^{-1} \propto \omega^4 \,, \tag{5.47}$$

$$\tau_{\mathrm{U}}^{-1} \propto T\omega^2 \exp\left(-\frac{\Theta}{bT}\right) \,, \tag{5.48}$$

where l is the characteristic size of the sample, $\Theta = \hbar\omega_{\mathrm{D}}/k$ is the Debye temperature, and b is a constant of the order of a few units. Note that other expressions have been put forward for τ_{U}^{-1} [13, 14].

Phonon scattering by particles can be treated classically. If we assume that the field scattered by a particle is not affected by the presence of the other particles (single scattering hypothesis), then

$$\tau_{\mathrm{P}}^{-1} = Nv\sigma \,, \tag{5.49}$$

where N is the particle concentration per unit volume, v the speed of sound, and σ the scattering cross-section by a single particle.

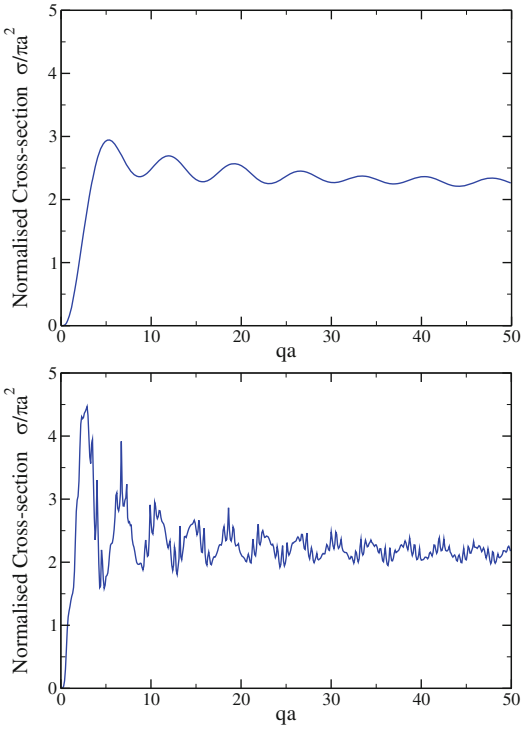

Fig. 5.5 Normalised scattering cross-section $\sigma/\pi a^2$ as a function of qa, for scattering of longitudinal waves, where a is the radius of the (spherical) particles, and q is the wave number. *Top*: Ge particles in an Si matrix. *Bottom*: Si particles in an Ge matrix

5.2.3 Phonon Scattering by Particles

The scattering cross-section σ of a single particle can be calculated using the theory of acoustics in continuous media. The case of a spherical fluid particle in a fluid medium was treated by Morse et al. [15]. This approach ignores transverse waves and is therefore in principle poorly suited to the study of solids. The scattering of acoustic waves, either longitudinal or transverse, by solid particles immersed in a solid elastic medium generates a scattered field with both polarisations. This problem has been treated analytically for isotropic matrices and particles [16–19]. Recently, molecular dynamics has been used to study particles with various shapes and sizes, immersed in an anisotropic elastic medium [20]. Moreover, this technique allows one to introduce a degree of roughness into the particle surface.

At low frequencies ($qa \ll 1$), the cross-section has a Rayleigh behaviour $\sigma \propto \omega^4$. At high frequencies ($qa \gg 1$), it tends to $2\pi a^2$, i.e., a geometric behaviour. Between these two regimes, there is no simple analytic expression for σ, and it evolves in an oscillatory manner, with an amplitude and period that depend on the relative elastic properties of the materials present. Figure 5.5 shows two example cross-sections for the scattering of longitudinal waves by spherical particles of radius a [16, 19]. The constituents are considered to be isotropic and the parameters used are average values for Si and Ge (see Appendix C).

In reality, heat conduction is not sensitive to the exact form of the cross-section. Indeed, integration over the phonon spectrum averages out any scattering minima or maxima. This is illustrated in Fig. 5.6. Here the conductivity has been calculated using two expressions for τ_P and applying (5.42) with a particle volume fraction of 1%. The first expression for τ_P corresponds to Fig. 5.5a, while the second takes $\sigma = 2\pi a^2$ over the whole frequency range. The results are shown in Fig. 5.6 by the continuous and dashed curves. Note that the discrepancy between the two curves is extremely slight between 10 K and 300 K. Over this temperature range, the main phonons carrying the heat ($\hbar\omega \sim kT$) satisfy $5 < qa < 500$ in the case considered ($a = 10$ nm, $v_D = 3\,600\ \mathrm{m\,s^{-1}}$). We thus find that, for this range of values of qa, the approximation $\sigma = 2\pi a^2$ gives good results.

Compared with the other scattering mechanisms, scattering by nanoparticles tends to be more efficient on low and medium frequency phonons. Their role is thus enhanced when heat is transported mainly by low frequency phonons. This is what happens naturally at low temperatures, but also at high temperatures if the role of high frequency phonons has been minimised by other processes. This is illustrated in the following two examples.

5.2.4 Example of a Pure Bulk Material

Consider a bulk material that is relatively pure, i.e., with a low level of impurities, containing a small amount of nanoparticles (volume fraction η_p typically less than or equal to 1%). This is illustrated in Fig. 5.6. Conductivities were calculated using

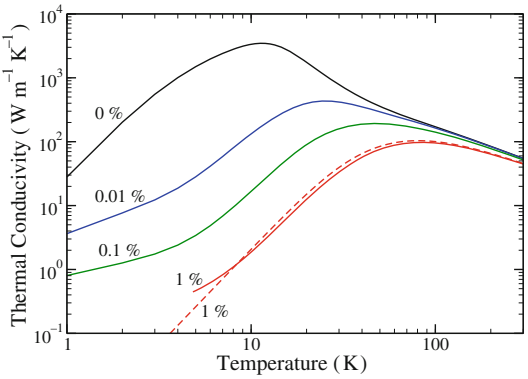

Fig. 5.6 Influence of particles on the thermal conductivity of a pure bulk material. Curves are parametrised by the particle volume fraction (see Appendix C). The *dashed curve* is the result calculated using $\sigma = 2\pi a^2$

(5.42). The parameters, given in Appendix C, are representative of a Ge matrix containing Si nanoparticles.

The effect of the particles is most noticeable at low temperatures, and Fig. 5.7 explains why this comes about. The figure shows the relaxation times associated with the various processes. Only the Umklapp processes introduce any temperature dependence. At high temperatures, a significant part of the energy is transported by very high frequency phonons, i.e., $\hbar\omega/kT \sim 1$–5, whence $\omega \sim 1$–6×10^{13} rad s^{-1} at 100 K. Figure 5.7 shows that, over this frequency range, relaxation times are dominated by Umklapp processes. The effect of the nanoparticles is thus very limited and the conductivity is barely sensitive to their presence. At low temperatures, the phonons carrying the heat are low frequency, i.e., $\hbar\omega/kT \sim 1$–5, whence

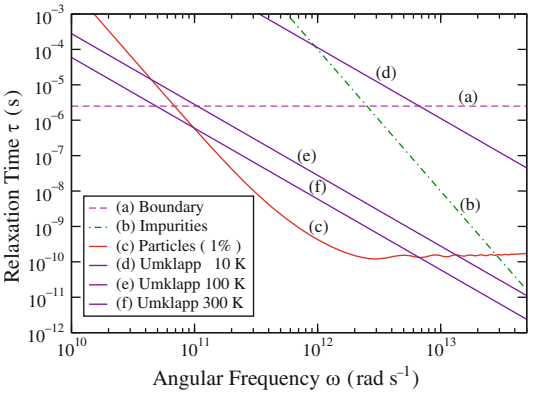

Fig. 5.7 Particles in a pure bulk material. Relaxation times for the different processes as a function of the phonon angular frequency (see Appendix C for the parameters)

$\omega \sim 1$–6×10^{12} rad s^{-1} at 10 K. Figure 5.7 shows that, over this frequency range, the nanoparticles are now responsible for the process dominating the relaxation time. As a result, their presence leads to a significant reduction in the conductivity at low temperatures.

This type of behaviour has been observed experimentally and interpreted in terms of scattering by nanoparticles. For example, in KBr and KCl, nanoscale precipitates ($\sim$ 10 nm) of $SrBr_2$, $SrCl_2$, and $BaBr_2$ lead to a decrease in conductivity at low temperatures [7]. In NaCl, 20 nm silver particles have a significant effect at low temperatures, and much less effect at higher temperatures [8].

We have been considering small volume fractions of particles here. For high concentrations, a significant reduction in conductivity is observed, even at high temperatures [10].

5.2.5 *Example of a Disordered Alloy*

When the impurity concentration starts to get higher, a minimum of thermal conductivity is reached for compositions corresponding to a disordered alloy. In this regime, at room temperature, scattering of high frequency phonons is maximal. Low frequency phonons then contribute significantly to transport, even at high temperatures. The inclusion of small amounts of nanoparticles (volume fraction of the order of 1%) can then scatter these low frequency phonons and thereby further reduce the heat conduction.

To illustrate this mechanism, we consider here a thin film ($\sim$ 1 μm) of highly doped material. Figure 5.8 shows the integrand $\tau(\omega,T)g(\omega,T)$ of (5.42) giving the conductivity. It gives the spectrum of the phonons dominating the transport for one

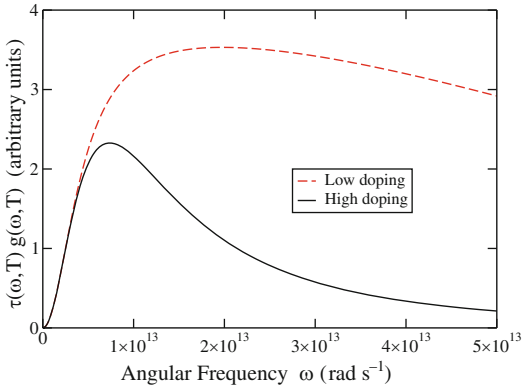

Fig. 5.8 Thin film. Effect of impurities on the spectrum of the dominant phonons. The function $\tau(\omega,T)g(\omega,T)$ of (5.42) is given for different values of ω at fixed $T = 300$ K (see Appendix C for the parameters)

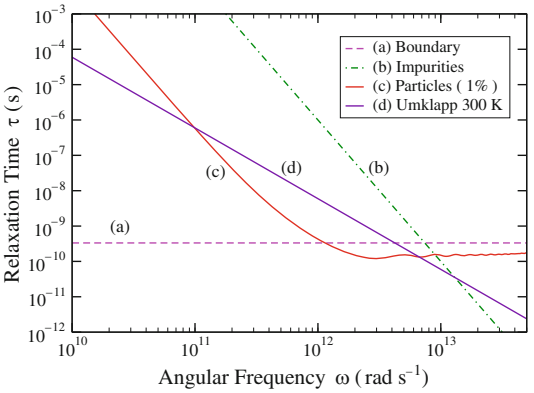

Fig. 5.9 Thin film. Relaxation times for the various processes as a function of the phonon angular frequency (see Appendix C for the parameters)

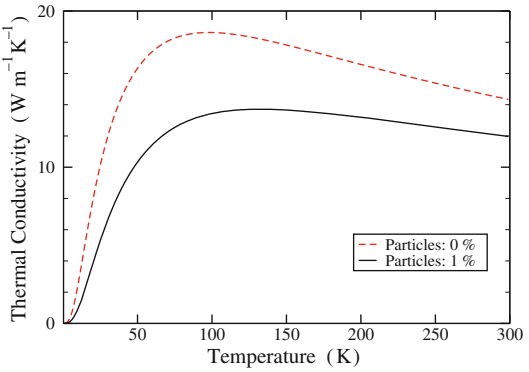

Fig. 5.10 Effect of nanoparticles on the thermal conductivity of a highly doped thin film (see Appendix C for the parameters)

low and one high concentration of impurities. It can be seen that the high doping eliminates the very high frequency phonons and shifts the spectrum towards lower frequencies. By introducing particles, one can then act on the low frequency part of the spectrum. The relaxation times are shown in Fig. 5.9, for high impurity doping, at 300 K. It can be seen that the addition of particles affects the low frequency phonons $\omega \sim 10^{12}$–10^{13} rad s^{-1}. This leads to a significant reduction in the thermal conductivity at room temperature, as can be seen from Fig. 5.10.

Kim et al. [11] measured the thermal conductivity of the alloy $In_xGa_{1-x}As$ for $x = 0.53$. This composition corresponds to the minimum conductivity of the alloy. They showed that, by introducing ErAs nanoparticles of diameter 1–4 nm in a volume fraction of the order of 1%, the conductivity could be reduced below the minimum. They interpreted their result by the scattering of low and medium frequency phonons [11, 12].

5.3 Conclusion

The appropriate method for handling the problem of heat conduction in composites depends on the size of the particles. The effective medium theories and methods treating the scattering of excitations by the particles are thus complementary. Effective medium theories account phenomenologically for the relaxation processes in the particles and matrix through the thermal conductivities of the constituents. They are well suited to the case of 'large' particles. Models for scattering by particles use a microscopic description for the heat transport mechanisms. Phonon scattering is treated in the framework of acoustics in isotropic continuous media. Absorption by the particles is not taken into account. These approaches are well suited to the case of 'small' particles.

It would be interesting to bring the two approaches closer together. For example, Minnich and Chen have suggested using the effective medium model to study the nanoparticle case [21]. To do this, they modify the thermal conductivities of the matrix and particles, arguing from the change in mean free path induced by nanostructuring the material. Conversely, the phonon scattering model could be suitable for dealing with large particles if it took into account relaxation processes within the particles.

Finally, the two approaches considered here simplify the problem of interparticle interaction and only apply to low particle concentrations. The effective medium model accounts for the interactions in an approximate way by assuming that the thermal field acting on the particle is a mean field related to the presence of the other particles. As far as the phonon scattering model is concerned, it assumes that each particle scatters an acoustic field that is independent of the field scattered by the other particles (single-particle scattering). Interactions between particles doubtless play an important role, and it would be useful to be able to handle it more precisely.

Appendix A. Demonstration of (5.7)

Let f be a scalar function and $\mathbf{V}$ a vector field. Then

$$\mathrm{div}(f\mathbf{V}) = f\mathrm{div}\mathbf{V} + (\mathrm{grad}f)\cdot\mathbf{V}\,. \tag{5.50}$$

For $f(\mathbf{r}) = z$ and $\mathbf{V} = \mathbf{q}$, where $\mathbf{r}(x,y,z)$ runs over the points of space and $\mathbf{q}$ is the local heat flux, we deduce that

$$\mathrm{div}(z\mathbf{q}) = z\,\mathrm{div}\mathbf{q} + q_z\,. \tag{5.51}$$

Furthermore, in the steady state,

$$\mathrm{div}\mathbf{q} = 0\,. \tag{5.52}$$

Hence,

$$\operatorname{div}(z\mathbf{q}) = q_z \,. \tag{5.53}$$

Since the Green–Ostrogradsky theorem implies that

$$\iiint_V \operatorname{div}(z\mathbf{q})\mathrm{d}v = \iint_{\partial V} z\mathbf{q}\cdot\mathrm{d}\mathbf{S} \,, \tag{5.54}$$

we have finally

$$\iiint_V q_z \mathrm{d}v = \iint_{\partial V} z\mathbf{q}\cdot\mathrm{d}\mathbf{S} \,, \tag{5.55}$$

which is (5.7).

Appendix B. Effective Medium and Interface Resistance

If a field g has no discontinuities,

$$\iiint_V \nabla g \mathrm{d}v = \iint_{\partial V} g\mathrm{d}\mathbf{S} \,. \tag{5.56}$$

The field T has discontinuities at the particle–matrix interfaces if there is any heat resistance at these interfaces. In this case (5.56) cannot be applied directly. In order to use (5.56), the field T is slightly modified and the discontinuities are replaced by continuous but fast temperature changes, occurring over a thin layer surrounding the particle.

Projecting onto the z direction and applying (5.56) to the modified T gives

$$\iiint_V \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v = \iint_{\partial V} T n_z \mathrm{d}s \,, \tag{5.57}$$

where n_z is the z component of the normal to the surface ∂V. Referring to Fig. 5.1,

$$\iint_{\partial V} T n_z \mathrm{d}s = (T_2 - T_1)S \,. \tag{5.58}$$

We now decompose the volume integral into a sum of integrals over the matrix m, the particle p, and the thin layer c:

$$\iiint_V \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v = \iiint_{\mathrm{m}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v + \iiint_{\mathrm{p}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v + \iiint_{\mathrm{c}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v \,. \tag{5.59}$$

The first two integrals on the right-hand side can be rewritten

$$\iiint_{\mathrm{m}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v = V_{\mathrm{m}} \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_{\mathrm{m}} \,, \qquad \iiint_{\mathrm{p}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v = V_{\mathrm{p}} \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_{\mathrm{p}} \,. \tag{5.60}$$

The third integral on the right-hand side of (5.59) is transformed using (5.57) to give

$$\iiint_{\mathrm{c}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v = \iint_{\partial \mathrm{c}} T n_z \mathrm{d}s \,, \tag{5.61}$$

where $\partial \mathrm{c}$ is the surface of the thin layer, comprising two disjoint surfaces, $\partial \mathrm{c}_{\mathrm{in}}$ and $\partial \mathrm{c}_{\mathrm{out}}$, whence

$$\iiint_{\mathrm{c}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v = \iint_{\partial \mathrm{c}_{\mathrm{in}}} T n_z \mathrm{d}s + \iint_{\partial \mathrm{c}_{\mathrm{out}}} T n_z \mathrm{d}s \,. \tag{5.62}$$

On the inner surface $\partial \mathrm{c}_{\mathrm{in}}$, the temperature T is evaluated on the particle side of the particle–matrix interface, while on the outer surface $\partial \mathrm{c}_{\mathrm{out}}$, it is evaluated on the matrix side of the interface. Finally, since the layer is very thin, the areas of the surfaces $\partial \mathrm{c}_{\mathrm{in}}$ and $\partial \mathrm{c}_{\mathrm{out}}$ are almost equal, while their normals are oriented directly opposite one another. It follows that

$$\iiint_{\mathrm{c}} \frac{\mathrm{d}T}{\mathrm{d}z}\mathrm{d}v = \iint_{I} (T_{\mathrm{m}} - T_{\mathrm{p}}) n_z \mathrm{d}s \,, \tag{5.63}$$

where I stands for the union of all particle–matrix interfaces throughout V. Therefore,

$$(T_2 - T_1)S = V_{\mathrm{m}} \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_{\mathrm{m}} + V_{\mathrm{p}} \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_{\mathrm{p}} - \iint_{I} (T_{\mathrm{p}} - T_{\mathrm{m}}) n_z \mathrm{d}s \,, \tag{5.64}$$

and dividing by the volume $V = SL$,

$$\frac{T_2 - T_1}{L} = \eta_{\mathrm{m}} \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_{\mathrm{m}} + \eta_{\mathrm{p}} \left\langle \frac{\mathrm{d}T}{\mathrm{d}z} \right\rangle_{\mathrm{p}} - \frac{\eta_{\mathrm{p}}}{V_{\mathrm{p}}} \iint_{I} (T_{\mathrm{p}} - T_{\mathrm{m}}) n_z \mathrm{d}s \,. \tag{5.65}$$

Appendix C. Calculation Parameters for Scattering by Particles

The aim here is to provide the calculation parameters used in Sect. 5.2. Scattering cross-sections σ are calculated for the scattering of longitudinal waves [16, 19]. In the context of the Debye model, it was then assumed that $\tau_{\mathrm{P}}^{-1} = N v_{\mathrm{D}} \sigma$.

Table 5.1 Characteristics of the materials: v_{L} and v_{T} are the speeds of the longitudinal and transverse waves, v_{D} is the Debye speed, ρ the density, Θ the Debye temperature, and a the particle radius

		Figure 5.5 top	Figure 5.5 bottom	Figures 5.6, 5.7, 5.9, and 5.10	Figure 5.8
Particles	v_{L} [m s^{-1}]	8 970	5 300	8 970	
	v_{T} [m s^{-1}]	5 200	3 110	5 200	
	ρ [kg m^{-3}]	2 330	5 320	2 330	
	a [nm]			10	
Matrix	v_{L} [m s^{-1}]	5 300	8 970	5 300	5 300
	v_{T} [m s^{-1}]	3 110	5 200	3 110	3 110
	v_{D} [m s^{-1}]			3 600	3 600
	ρ [kg m^{-3}]	5 320	2 330	5 320	5 320
	Θ [K]			400	400

Table 5.2 Characteristics of the scattering processes

		Figures 5.6, 5.7	Figure 5.8	Figures 5.9, 5.10
Boundary $\tau_{\mathrm{B}}^{-1} = C$	C [s^{-1}]	4×10^5	3×10^9	3×10^9
Umklapp	B [s K^{-1}]	7×10^{-19}	7×10^{-19}	7×10^{-19}
$\tau_{\mathrm{U}}^{-1} = BT\omega^2 \mathrm{e}^{-\Theta/bT}$	b	6	6	6
Impurities	A [s^3]	10^{-44}	10^{-44}	
$\tau_{\mathrm{I}}^{-1} = A\omega^4$			10^{-42}	10^{-42}
Particles $\tau_{\mathrm{P}}^{-1} = N v_{\mathrm{D}} \sigma$	N [m^{-3}]	η_{P} N	η_{P} N	η_{P} N
	η_{P} [%]	0 0	0 0	0 0
		0.01 2.4×10^{19}		
		0.1 2.4×10^{20}		
		1.0 2.4×10^{21}		1.0 2.4×10^{21}

References

1. Y. Benveniste: Effective thermal conductivity of composites with a thermal contact resistance between the constituents: Nondilute case, J. Appl. Phys. **61**, 2840 (1987)
2. C.-W. Nan, R. Birringer, D.R. Clarke, and H. Gleiter: Effective thermal conductivity of particulate composites with interfacial thermal resistance, J. Appl. Phys. **81**, 6692 (1997)
3. G. Every, A.Y. Tzou, D.P.H. Hasselman, and R. Raj: The effect of particle size on the thermal conductivity of ZnS/diamond composites, Acta Metall. Mater. **40**, 123 (1992)
4. D.P.H. Hasselman, K.Y. Donaldson, and A.L. Geiger: Effect of reinforcement particle size on the thermal conductivity of a particulate-silicon carbide-reinforced aluminum matrix composite, J. Am. Ceram. Soc. **75**, 3137 (1992)
5. S.A. Putnam, D.G. Cahill, B.J. Ash, and L.S. Schadler: High-precision thermal conductivity measurements as a probe of polymer/nanoparticle interfaces, J. Appl. Phys. **94**, 6785 (2003)
6. P.G. Klemens: *Solid State Physics* (see the chapter entitled Thermal Conductivity and Lattice Vibrational Modes), Academic Press Inc., New York (1958)
7. J.W. Schwartz and C.T. Walker: Thermal conductivity of some alkali halides containing divalent impurities. (ii) Precipitate scattering. Phys. Rev. **155** (3), 969–979 (1967)
8. J.M. Worlock: Thermal conductivity in sodium chloride crystals containing silver colloids, Phys. Rev. **147**, 636 (1966)
9. A. Khitun, A. Balandin, J.L. Liu, and K.L. Wang: In-plane lattice thermal conductivity of a quantum-dot superlattice, J. Appl. Phys. **88**, 696 (2000)
10. J.L. Liu, A. Khitun, K.L. Wang, W.L. Liu, G. Chen, Q.H. Xie, and S.G. Thomas: Cross-plane thermal conductivity of self-assembled Ge quantum dot superlattices, Phys. Rev. B **67** (16), 165333 (2003)
11. W. Kim, J. Zide, A. Gossard, D. Klenov, S. Stemmer, A. Shakouri, and A. Majumdar: Thermal conductivity reduction and thermoelectric figure of merit increase by embedding nanoparticles in crystalline semiconductors, Phys. Rev. Lett. **96**, 045901 (2006)
12. W. Kim and A. Majumdar: Phonon scattering cross-section of polydispersed spherical nanoparticles, J. Appl. Phys. **99**, 084306 (2006)
13. M.G. Holland: Analysis of lattice thermal conductivity, Phys. Rev. **132** (6), 2461–2471 (1963)
14. J.W. Schwartz and C.T. Walker: Thermal conductivity of some alkali halides containing divalent impurities. (i) Phonon resonances, Phys. Rev. **155** (3), 959–969 (1967)
15. P.M. Morse and K.U. Ingard: *Theoretical Acoustics*, Princeton University Press, Princeton, New Jersey (1986)
16. C.L. Ying and R. Truell: Scattering of a plane longitudinal wave by a spherical obstacle in an isotropically elastic solid, J. Appl. Phys. **27**, 1086 (1956)
17. N.G. Einspruch, E.J. Witterholt, and R. Truell: Scattering of a plane transverse wave by a spherical obstacle in an elastic medium, J. Appl. Phys. **31**, 806 (1960)
18. R.J. McBride and D.W. Kraft: Scattering of a transverse elastic wave by an elastic sphere in a solid medium, J. Appl. Phys. **43**, 4853 (1972)
19. R. Truell, C. Elbaum, and B.B. Chick: *Ultrasonic Methods in Solid State Physics*, Academic Press, New York and London (1969)
20. N. Zuckerman and J.R. Lukes: Acoustic phonon scattering from particles embedded in an anisotropic medium: A molecular dynamics study, Phys. Rev. B **77**, 094302 (2008)
21. A. Minnich and G. Chen: Modified effective medium formulation for the thermal conductivity of nanocomposites, Appl. Phys. Lett. **91**, 073105 (2007)

Chapter 6
Optical Generation and Detection of Heat Exchanges in Metal–Dielectric Nanocomposites

Bruno Palpant

In this chapter, we shall be concerned with the thermal properties of nanocomposite materials made up of metal nanoparticles dispersed throughout a transparent host medium, and in particular, the role they play in the optical response of such materials. Indeed, these two types of phenomenon are closely related, we shall show how to exploit this relationship by using the metal nanoparticles both as a nanoscale heat source and as a probe for local temperature variations, via their optical properties. The relationship between optical and thermal response in these media can be demonstrated today using ultrafast optical techniques to study their relaxation dynamics. For this reason we shall consider in particular the time dependence of heat exchanges in nanocomposites.

Apart from its application to simulating and predicting the optical response of nanocomposite media, the determination of temperature dynamics is interesting in itself from the point of view of the energy conversion effected by metal nanoparticles. This is exemplified by devices for plasmonics [1], where local heating can alter the way electromagnetic waves are guided in the near field by metal nano-objects, an effect that needs to be carefully controlled, or by labelling for microscopy in biology, where nanoparticles are heated in order to modify the optical response of their local environment [2]. We should also mention the method of thermal therapy proposed recently for cancer treatment [3, 4]. Here, clusters of metal nanoparticles absorb light energy transmitted through the body tissues and convert it into heat, which then diffuses locally into the surrounding medium. If the particles have been preferentially distributed in the vicinity of sick cells by some kind of addressing mechanism, those cells will be destroyed by local heating. Another example is provided by optical limitation devices in which it has been shown that the modification of the local environment by metal nanoparticles, e.g., formation of gas bubbles, when they are heated by a laser pulse may underlie the limitation phenomenon [5,6]. Following the same line of thought, metal nanoparticles are considered as model defects for studying the deterioration of optical elements under pulsed laser radiation [7–9]. Once again the dynamics of the light–heat conversion in a nanoparticle and

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 127
DOI 10.1007/978-3-642-04258-4_6, © Springer-Verlag Berlin Heidelberg 2009

subsequent heat transfer into the local environment are crucial for understanding damage mechanisms.

This chapter has three sections intended to provide the reader with the basic tools for understanding the generation and detection of heat exchanges in metal nanoparticles via their optical response. We begin with a summary of the key features concerning the various contributions to this response. We then consider the question of temperature dependence, and end by modelling the dynamics of heat exchanges. The term 'thermal effects' will be taken in a broad sense, encompassing not only the non-equilibrium gas of conduction electrons, but also heat exchanges within nanocomposite media.

6.1 Optical Properties of Noble Metal Nanoparticles and Nanocomposite Media

The linear propagation of an electromagnetic wave in a homogeneous and isotropic medium is governed by the usual complex optical index $\tilde{n} = n + \mathrm{i}\kappa$ of the medium. Here n is the refractive index and κ the extinction coefficient, proportional to the absorption coefficient $\alpha = 4\pi\kappa/\lambda$, where λ is the wavelength of the incident radiation. The complex index is related to the dielectric function $\varepsilon = \varepsilon_1 + \mathrm{i}\varepsilon_2$ by $\varepsilon = \tilde{n}^2$. We shall now turn to the optical properties of the noble metals in their bulk phase.

6.1.1 Dielectric Function of Noble Metals

The dielectric function of the noble metals – copper, silver, and gold – has been the subject of many experimental investigations [10–15]. A review and analysis of the main results can be found in [16]. In contrast to the situation with the alkali metals, the response of the noble metals to an electromagnetic excitation in the UV–visible range cannot be described solely in terms of the behaviour of the quasi-free conduction electrons (*sp* band), but must include the influence of bound electrons in the *d* bands. As a consequence, the total dielectric function ε_m of the noble metals can be written as the sum of two contributions, one due to electron transitions in the conduction band (intraband transitions) and the other arising from transitions from the *d* bands to the conduction band (interband transitions), viz., $\varepsilon_\mathrm{m} = \chi^\mathrm{f} + \varepsilon^\mathrm{ib}$.

Intraband Contribution: Drude Model

The contribution χ^f of quasi-free electrons to the dielectric function can be described classically via the Drude model [17], in which the basic ingredients are the bulk plasmon angular frequency ω_p and a phenomenological damping constant Γ accounting for all the scattering processes suffered by the conduction electrons of

the metal. The scattering processes are assumed independent and the values of the corresponding Γ constants are therefore additive (Matthiessen's rule [18]). It can be shown that, for electron temperatures up to a few thousand kelvin, the dominant contribution to Γ comes from electron–phonon collisions. For a nanoparticle, another scattering channel arises when the size of the particle is less than the electron mean free path. One must then add a scattering term that is inversely proportional to the particle size, where the factor of proportionality may depend on the frequency of the exciting wave [19].

Interband Contribution: Rosei Model

By virtue of the Pauli exclusion principle, there is a minimal photon energy for which the interband (ib) transition can occur, corresponding to the excitation of an electron from the top of the valence band to the Fermi level (Fermi energy E_F) situated in the conduction band. This defines an energy threshold below which the imaginary part of ε_{ib} is zero. Whereas this threshold falls in the ultraviolet (UV) for silver (≈ 3.9 eV) [10], it lies in the visible for gold and copper (at around 2.4 and 2.1 eV, respectively) [10, 13]. This explains the colours of these metals in the bulk phase. Unlike the intraband transitions, the interband contribution ε^{ib} to the dielectric function of the noble metals cannot be calculated classically. One must then evaluate the sum over all possible transitions from an occupied state of the valence band to a free state in the conduction band. A detailed model of the band structure of the metal is then required [10, 20–23]. This problem was tackled in the 1970s by Rosei and coworkers [21–24].

Calculation Procedure

In general, only the imaginary part of ε^{ib} is determined. The real part is then deduced using the Kramers–Kronig analysis. It can be shown within the framework of the Lindhard theory that the product $\omega^2\varepsilon_2^{ib}$ is proportional to the integral of the product of the transition probability and the difference in occupation numbers f of the levels involved in the transition, with the integral taken over all possible transitions and all possible wave vectors $\mathbf{k}$ of the Brillouin zone (BZ) [17, 22]. For the three noble metals, the interband transitions involved in the optical domain occur between occupied levels of the d bands and vacant levels in the hybridised sp band. The X point (threshold ≈ 1.9 eV in gold [13, 24]) and especially the L point (threshold ≈ 2.4 eV in gold [24]) of the BZ contribute mainly to this type of transition in the visible. In the vicinity of each of these two points, it can be shown that ε_2^{ib} is proportional to the joint density of states, i.e., associated with a transition of given energy, which is equal to the density of states in the d and sp bands satisfying energy and wave vector conservation [21].

Since the d band states are completely filled, the problem simplifies because $f_d = 1$ for any $\mathbf{k}$. Moreover, the electron density f_{sp} in the conduction band can be

found by solving Boltzmann's equation (see p. 135 in Sect. 6.2.3). By replacing the integration over the wave vectors by an equivalent energy integral, one can then calculate the density of states, which depends on the structure parameters of the bands (energies E_d and E_{sp} of the bands, effective electron mass m_{eff}) near the relevant points of the BZ.

As we shall see in Sect. 6.2.3, this approach is commonly used to calculate the modification of the optical properties of noble metal nanoparticles induced by absorption of an ultrashort laser pulse. For gold, the contribution of transitions in the neighbourhood of the L point then dominates.

Band Structure

The band structures of gold and silver were studied at the beginning of the 1970s [25]. The *d* bands are located a few eV below the Fermi level. The *sp* band crosses the latter near the X, L, and Σ points of the BZ. Locally, in a given direction of the reciprocal (momentum) space, it displays the characteristics of quasi-free electron behaviour (parabolic band). It was this observation that led Rosei and coworkers [24] to model the band structure near the points making the main contribution to the interband optical response, viz., the X and L points, by a set of parabolic branches, with a dispersion law of the form $E(\mathbf{k}) = \hbar\mathbf{k}^2/2m_{\text{eff}}$, as illustrated in Fig. 6.1 in the vicinity of the L point for gold.

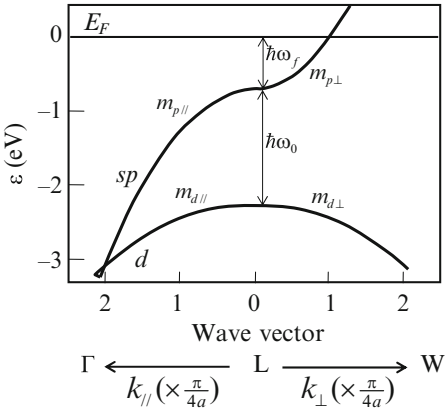

Fig. 6.1 Model for the band structure of gold near the L point of the Brillouin zone and close to the Fermi level. Taken from [24, 25]. Bands are described locally by parabolic branches. The parameters arising in the calculation of the joint density of states are indicated. The scale of the wave vectors is in units of $\pi/4a$, where a is the lattice parameter of gold

6.1.2 Optical Response of Nanocomposite Media

Surface Plasmon Resonance in Nanoparticles

When a metal is divided into small pieces, electron confinement can change its optical response in a spectacular way. Indeed, if the size of these pieces is very small compared with the wavelength of the electromagnetic radiation, all the conduction electrons feel the same homogeneous field and oscillate collectively, like a giant dipole. The coupling is resonant when the frequency of the wave is tuned to the natural frequency of the oscillations of the electron gas around the positive ionic core. This phenomenon is known as surface plasmon resonance (SPR), since the excess charges appear at the particle surface during the oscillation, by analogy with the volume plasmon occurring in the bulk phase. From the standpoint of quantum theory, the SPR corresponds to the coherent excitation of electron transitions in the conduction band. It is also exhibited by calculating the extinction cross-section using Mie theory to dipole order [26]. At resonance, the strong polarisation in the metal induces a local electromagnetic field within and in the neighbourhood of the particle, whose amplitude may exceed that of the applied field. This property has been exploited in many technological developments in the fields of nonlinear optics [27], sensors, biomedical imaging, and so on.

In silver, the SPR, occurring in the UV, has a greater oscillating strength than the SPR in gold, or indeed copper, which are situated in the visible. This is due to the significant spectral coupling in gold or copper between core electrons and conduction electrons. Moreover, the SPR is all the more pronounced the higher the refractive index of the host medium. Finally, the SPR shifts toward the red as the refractive index increases. As an example, it is located around 2.33 eV for Au:SiO_2, 2.94 eV for Ag:SiO_2, and 2.29 eV for Au:Al_2O_3. Several general works discuss in detail the SPR in noble metal nanoparticles and nanocomposites [28–30].

From a thermal standpoint, it is easy to understand the interest of the SPR. For one thing, this phenomenon provides an efficient way of injecting a large amount of energy into nanoscale objects via electromagnetic illumination, in a spectral range that is perfectly accessible to standard lasers. For another, the significant influence of the SPR on the optical response of a nanocomposite material can reveal modifications in the latter induced by heat exchange processes within the relevant media. Both aspects are exploited in a pump–probe experiment. A high energy pulse (pump) perturbs the material, and a second, low energy pulse is subsequently used to probe this perturbation. By measuring the intensity of the probe beam and varying the delay between the two pulses, the relaxation dynamics of the medium can be reconstituted.

Nanocomposite Materials

In a medium containing a high density of nanoparticles, ensemble effects can no longer be neglected. In particular, mean field effects and electromagnetic interactions

between neighbouring nano-objects cannot be ignored. Among the various approaches developed to include these effects when calculating the optical response of nanocomposite materials, the effective medium theories are the most widely used. They are suited to a specific morphology and range of concentrations, and they can in general be used to obtain an analytic expression for the effective dielectric function of the material. The first, best known, and simplest of these is the Maxwell-Garnett theory [31], which assumes zero interactions between particles. Quite generally, the characteristics of the absorption band related to the SPR (position, intensity, width) change with the metal concentration p. Other methods, based on numerical solution of the electromagnetic equations governing the coupling of a wave with a virtual sample of the inhomogeneous medium, can provide an interesting alternative to effective medium theories.

6.2 Thermo-Optical Response

We now discuss changes in the optical properties of nanocomposite media induced by changes in their temperature.

6.2.1 Noble Metals

The thermo-optical response of noble metals has several origins, and these were the subject of many investigations in the 1970s, particularly concerning gold [21,24,32–35]. Various mechanisms can lead to a modification in the interband and intraband transitions following a temperature variation, altering the electron distribution directly, influencing lattice properties by increasing the interatomic distance, or affecting scattering phenomena. The main effects are shown schematically in Fig. 6.2.

Phenomenologically, the change in the complex index can be related linearly, to first order, to the change in temperature, making use of the thermo-optical coefficients $\mathrm{d}_T n = \partial n/\partial T$ and $\mathrm{d}_T \kappa = \partial \kappa/\partial T$:

$$\Delta \tilde{n} = (\mathrm{d}_T n + \mathrm{i}\mathrm{d}_T \kappa)\Delta T \ . \tag{6.1}$$

For example, the thermo-optical coefficients $\mathrm{d}_T n$ and $\mathrm{d}_T \kappa$ for gold, averaged over the range 295–670 K, have been determined by accurate analysis of results appearing in the literature, in particular by cross-checking the various results of spectroscopic thermomodulation experiments [36].

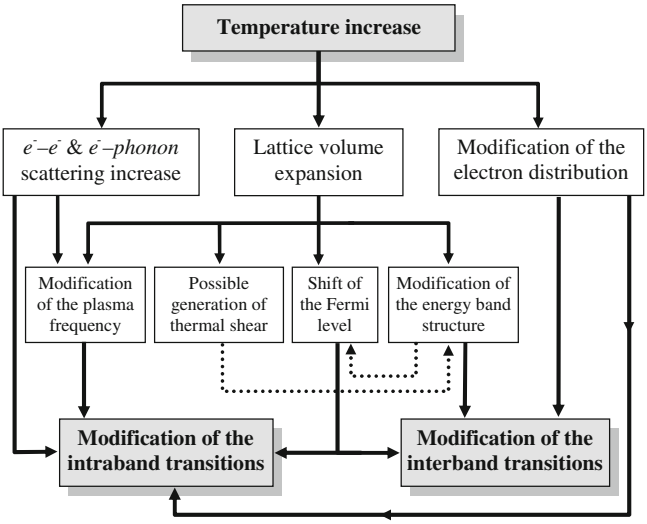
Temperature increase
e⁻–e⁻ & e⁻–phonon scattering increase
Lattice volume expansion
Modification of the electron distribution
Modification of the plasma frequency
Possible generation of thermal shear
Shift of the Fermi level
Modification of the energy band structure
Modification of the intraband transitions
Modification of the interband transitions

Fig. 6.2 Flow chart for the mechanisms affecting the thermo-optical response of the noble metals when the temperature increases. Taken from [36]

6.2.2 Nanocomposites

The thermo-optical response of a nanocomposite medium naturally depends on that of its constituents. Like the third-order nonlinear optical response, the thermo-optical response of a nanocomposite does not reduce to a simple average of the responses of its various components. Indeed, the local field factor, whose complex value varies significantly near the SPR, plays a key role.

Insofar as the optical properties of an inhomogeneous medium can be described by an effective medium approach, provided that the necessary assumptions are satisfied, such an approach can be extended to the temperature dependent case. To do this, one identifies the T dependent parameters in the expression for the effective susceptibility, which is then differentiated. Figure 6.3 shows the results of the calculation based on the extension of the Maxwell-Garnett theory (MGT) for Au:SiO_2 with $p = 8\%$. The temperature dependent quantities are the index of the matrix, the dielectric function of the metal, and the concentration of the metal (due to the different thermal expansion coefficients of the two types of material).

It can be shown that the effective thermo-optical response in the spectral range of the SPR is resonant. The resonance condition is exactly like the resonance condition of the optical response at the SPR. The metal particles have their own thermo-optical response, sometimes affected by interactions with their neighbours. The local field then plays the role of a complex magnifying glass, amplifying and distorting the spectrum of this intrinsic response about the plasmon resonance. This effect is particularly sensitive for gold nanospheres, where the SPR (in typical dielectrics) and the threshold of interband transitions, the main cause of modifications to the optical

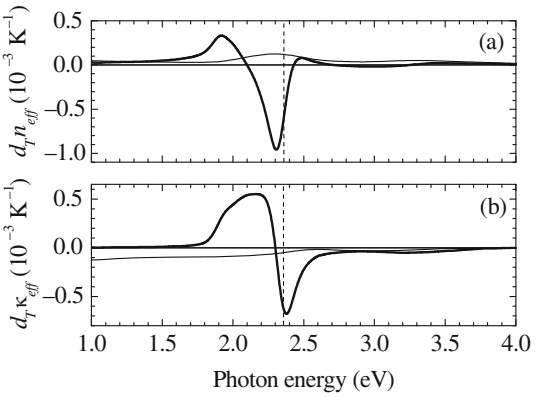

Fig. 6.3 Spectra of $d_T n_{\text{eff}}$ (**a**) and $d_T \kappa_{\text{eff}}$ (**b**) (*thick curves*) for an Au:SiO_2 nanocomposite with a concentration of 8%, derived using the MGT. The spectra corresponding to the volume average of the coefficients for gold and silica are also shown (*thin curves*). *Vertical dashed lines* indicate the position of the SPR

response, occur at very close photon energies. At first glance, the consequences are surprising: the thermo-optical coefficients of the medium as a whole are amplified by the local field enhancement effect, and display spectral variations with very different sign and amplitude to those observed for the various constituents taken separately or for a fictional medium whose properties were given by the volume average of those of its constituents (see Fig. 6.3). Temperature-dependent ellipsometry measurements confirm these results [37].

6.2.3 Calculation

The calculation of temperature variations in a nanocomposite following the absorption of a laser pulse will be discussed in Sect. 6.3. To begin with, we shall briefly describe in this section the way these temperature variations can be related to changes induced in the optical response. The idea here is to model the mechanisms discussed in Sect. 6.2.1, and then calculate their consequences. Depending on the case (time scales of excitation and observation, spectral range, excitation power, etc.), one can choose to focus on only a selection of these mechanisms, either the most important among them or those whose effects one is specifically interested in, or one can treat all of them at once.

Here we shall present the two commonest cases, defined by whether or not thermodynamic equilibrium has been established between electrons, metal lattice, and local environment (on the scale of one or a few nanoparticles).

Steady-State Regime

This case is rather easy to deal with, since it suffices to specify a single temperature in order to describe the thermodynamic state of a particle and its immediate neighbourhood. It is encountered when applying a continuous-wave laser in the steady-state regime or a pulsed laser with long pulse width compared with the characteristic time for nanoscale heat exchanges. Apart from heating by light absorption, this case also corresponds to thermomodulation experiments, or simply to situations where the temperature of the medium as a whole varies macroscopically.

One can then calculate the change in the complex index of the composite medium induced by the change in its temperature through its thermo-optical coefficients as defined above. When the temperature field within the medium has a spatial dependence on distance scales that are large compared with the characteristic dimensions of the nanostructures, the index field can be established in the same way. This situation typically corresponds to the application of a laser beam with non-uniform transverse temperature profile (this is the case for most beams), inducing a generalised thermal lensing phenomenon (modification of beam convergence and/or intensity). This effect has been demonstrated in nanocomposites by several groups, using nonlinear optical measurement results [38–43].

Transient Regime

The heating of a nanoparticle, or indeed a nanocomposite, has important consequences for its steady-state optical response, so one should only expect to find that it has a significant effect on the dynamics of this response in the transient regime. This concerns in particular the case where the medium is excited by an ultrashort light pulse, as happens in a pump–probe experiment or certain nonlinear optical setups. In the steady state, these consequences were grouped together under the generic heading of the thermo-optical response. Indeed, in that context, there is no need to treat the different physical causes separately (see Fig. 6.2). However, there is no avoiding this step if we wish to describe the optical response in the transient regime, since each mechanism acts differently and obeys its own dynamics. In Sect. 6.3 we shall see how to model the whole set of these mechanisms with a view to determining the time dependence of the temperature field (or internal energies) in a nanocomposite under pulsed laser excitation.

The optical response of a metal in the visible, near UV, or near IR is governed by the properties of the quasi-free conduction electrons and bound core electrons, as discussed in Sect. 6.1.1. The interband and intraband transition spectrum is determined by giving the distribution function $f(E,t)$ of the conduction electrons around the Fermi level, i.e., the probability of occupation of an electron level of energy E at time t. The function f is thus a relevant parameter for relating the microscopic behaviour of the metal to its optical response. When the electron gas is in an internal equilibrium state, i.e., when f can be described by Fermi–Dirac statistics, one can equally well use the electron temperature T_e.

For an ultrashort excitation, one sets up a model to describe the dynamics of the electron distribution (or T_e) in the metal subject to a light pulse, a point to be discussed in the next section. Once the time dependence of f has been determined using this model and the relevant input data (energy and duration of the pulse, wavelength, linear absorption coefficient, initial temperature, particle size, thermal conductivity of the host dielectric, metal concentration, and so on), the Drude model (see p. 128 of Sect. 6.1.1) and the Rosei approach (see p. 129 of Sect. 6.1.1) are used to calculate the change in the dielectric function of the metal. Finally, one appeals to an effective medium theory or any other method for carrying out electromagnetic calculations in inhomogeneous media in order to obtain the overall response of the nanocomposite. For example, one can simulate the differential reflection and transmission of a sample in a pump–probe experiment, or determine its nonlinear optical susceptibility.

6.3 Heat Exchange Dynamics

The optical response of a nanocomposite depends on various mechanisms, each with its own dynamics [44–54]: absorption of light energy by the electrons, redistribution within the gas of conduction electrons via electron–electron collisions, relaxation to the metal lattice by electron–phonon collisions, and finally, cooling of the particles by heat transfer into the host medium. During the first stage, the fast injection of energy into the electron gas shifts it out of equilibrium and induces a change in the transition spectrum involving the energy levels close to the Fermi level, and hence also modifies the optical response close to the IB threshold. The return to thermodynamic equilibrium of the whole medium through all the other processes is accompanied by a variation in the optical properties due to cooling of the electron gas, the change in temperature of the lattice, heating of the host medium, and heat exchange between neighbouring nanoparticles. Quite generally then, the mechanisms involved in the dynamics of the optical response have a largely thermal origin, in the broad sense of the term. Note that the very earliest stages of relaxation, viz., the first few picoseconds after the excitation, have been studied in detail by several groups [44, 47, 50, 52]. We shall describe only the basic features of the calculation, without entering into the various refinements.

6.3.1 Athermal Regime and the Boltzmann Equation

When a metal at temperature T_0 absorbs an ultrashort light pulse via transitions in the gas of conduction electrons,[1] the electron distribution is shifted out of equilibrium.

[1] For reasons of simplicity, we consider an excitation at a photon energy below the interband transition threshold. If this were not the case, we would have to include Landau damping, i.e., coupling of the plasmon with individual excitations.

Once the excitation has ended, internal thermalisation by electron–electron scattering allows it to recover a Fermi–Dirac profile at a temperature $T_e > T_0$, while electron–phonon collisions convert the energy into heat within the particle, whose temperature T_l thereby increases.[2] The phase during which the electron distribution is out of internal equilibrium is called the athermal regime. This phase lasts for a few hundred femtoseconds. It does not need to be considered for slow excitation conditions, where electrons and phonons are in thermodynamic equilibrium at any given time.

Boltzmann Equation for the Electron Distribution

In this regime, the relevant quantity for describing the properties of the metal is the distribution function $f(E,t)$, as explained above. Its time variation is governed by the Boltzmann equation. Neglecting electron scattering[3] and the influence of the matrix in the case of metal nanoparticles,[4] this can be written in the form

$$\frac{\partial f(E,t)}{\partial t} = \left.\frac{\partial f(E,t)}{\partial t}\right|_{\text{source}} + \left.\frac{\partial f(E,t)}{\partial t}\right|_{\text{e–e}} + \left.\frac{\partial f(E,t)}{\partial t}\right|_{\text{e–ph}} . \tag{6.2}$$

On the right-hand side, the first term describes the variation of f due to absorption of the light pulse (source term), while the other two refer to the variation rates of f due to electron–electron and electron–phonon scattering, respectively.

Over the last few years, several groups have developed and applied methods for solving the Boltzmann equation with varying degrees of accuracy, with the aim of describing the athermal regime in noble metals [55–61]. They range from the relaxation time approximation for the two scattering terms labelled e–e and e–ph in (6.2) to a complete description of these terms through the various scattering mechanisms involved. Here we shall outline one of these methods, which has the advantage of providing a reasonable model but at an acceptable cost in terms of computation time [62]. This approach, which is similar in some ways to a method used earlier by several authors [59–61], is based on the relaxation time approximation for e–ph scattering and the Landau theory of Fermi liquids for e–e scattering.

[2] Note that, since the notion of temperature is not strictly well defined in the statistical sense in a metal object with smaller dimensions than the mean free path of the heat carriers, T_l must be considered here rather as a measure of the internal thermal energy.

[3] Because the particles are small compared with the penetration depth of the wave, the excitation can be treated as uniform over the whole electron gas at each instant of time.

[4] This approximation remains valid throughout the athermal regime, provided that there is no chemical interaction at the interface of the metal and host medium.

Source Term

Dependent on the time and the electron energy $\varepsilon = E - E_F$, this term is proportional to the instantaneous power absorbed per unit metal volume, denoted by $P_{abs}(t)$, and to the instantaneous variation in the occupation number at given ε. This second factor is found by calculating the probability that an electron is raised to the energy ε, by absorbing a photon of energy $\hbar\omega$, minus the probability that an electron moves up to the $\varepsilon + \hbar\omega$ energy level. The coefficient of proportionality is then determined by requiring conservation of the total energy.

Electron–Electron Scattering Term

The term in the Boltzmann equation corresponding to electron–electron scattering comes from the disappearance of electrons with an excess energy ε and creation of electrons at this energy, following screened Coulomb scattering from other electrons. For weak perturbations, the Landau theory of Fermi liquids allows one to apply the relaxation time approximation to describe the first contribution [63]. This electron lifetime due to e–e collisions depends on ε via $\tau_{e-e} = \tau_0 E_F^2/\varepsilon^2$, a formula which reflects the fact that the probability of an electron scattering falls off as it gets closer to the Fermi level, which is in turn a direct consequence of the Pauli exclusion principle.[5] Here τ_0 is the lifetime the electrons would have if the exclusion principle did not apply. Rather than try to calculate this by some kind of approximation, it can be treated as an adjustable parameter of the model and then determined by comparing with the results of relaxation dynamics measurements reported in the literature. Its value is found to lie in the range 0.3–1.0 fs. The second contribution to e–e scattering is calculated from the quantities already defined, using the expression established by Ritchie [65].

Electron–Phonon Scattering Term

The third term on the right-hand side of (6.2) comes from the spontaneous emission of phonons, stimulated emission of phonons, and the absorption of phonons. In the last two cases, in contrast to the first, the scattering rate is proportional to the number of available states in the reservoir (the population of phonons in the reservoir being given by a Bose–Einstein distribution at temperature T_l). The e–ph scattering term is then split into two contributions, each treated using the relaxation time approximation. Both depend on the rate of energy transfer $\dot{q}$ from the electron gas to the phonon bath.[6] The second also depends on the number of phonons created up to time t by the excitation and the average energy of the phonons (given by the

[5] Note that τ_{e-e} should not be confused with the characteristic time for energy redistribution within the electron gas, in the range 300–500 fs for the noble metals [64].

[6] The electron–phonon energy relaxation time is of the order of 1 ps in the noble metals [5].

Debye model for the density of states). A dimensionless factor S describes the relative weight of the two processes of emission and absorption.

Implementation and Application

The Boltzmann equation can then be solved by a finite difference method with chosen energy and time steps. $\dot{q}$ and S are, like τ_0, free parameters to be fitted to experimental results. This fitting is made possible by the fact that these three parameters are independent of ε and t, but also because they affect the dynamics in quite distinct time domains. They can be determined by fitting the calculations to experimental data reported in the literature [19, 56].

Finite-Size Effects

The characteristics of the scattering processes governing the time dependence of the electron distribution can be modified by finite-size effects when the metal is present in nanoscale pieces. There is a global reduction in the relaxation times due to e–e and e–ph collisions when R decreases [5, 66, 67]. One of the causes for this behaviour lies in the reduced screening of electron–electron and electron–ion Coulomb interactions. The increase in the e–ph scattering rate can also be attributed to modification of the phonon spectrum of the metal due to the appearance of vibration modes intrinsic to confinement [68]. The result of these finite-size effects is a modification of the parameters τ_0 and $\dot{q}$ which can be included in the model described above.

Conclusion Concerning the Athermal Electron Regime

The athermal regime for the electron distribution, together with its consequences for the short-time dynamics of the optical response on the occasion of ultrafast laser excitation, can indeed by adequately accounted for by solving the Boltzmann equation. However, the model used here is only valid insofar as heat exchanges with the environment of the nanoparticles remain negligible. Once the gas of conduction electrons has reached internal equilibrium, i.e., when it can be described by a Fermi–Dirac distribution at temperature T_e, the system enters the thermal regime. The influence of the matrix can then be included using a three-temperature model.

6.3.2 Thermal Regime and the Three-Temperature Model

In many publications, only exchanges between electrons and phonons within the metal particle (two-temperature model) are taken into account to simulate the

ultrafast dynamics of the optical response. Now the presence of the matrix plays a significant role after only the first few picoseconds following the excitation, and it can then no longer be neglected. Furthermore, the two-temperature model provides no way of handling excitation by long pulses. With the model discussed here, on the other hand, it should be possible in principle to calculate the temperature variations, whatever pulse duration is considered.

Assumptions

To simplify the problem, we shall assume that the conduction electron gas thermalises instantaneously, i.e., we shall neglect the finite duration of the athermal regime. This hypothesis is justified for an ultrashort pulse excitation of high energy, for which this regime is indeed very short-lived (the injection of a large amount of energy into the electron gas favours fast internal redistribution), and of course for long pulses, for which the electron gas is at all times in thermodynamic equilibrium with the ionic lattice of the metal. However, the approaches developed for the two regimes can be matched up using a suitably modified three-temperature model [62], but we shall not discuss that here.

In the three-temperature model, we shall take the thermodynamic characteristics of the various media to be those of their bulk phase. It would be relatively easy to include certain effects related to confinement in a phenomenological manner. The metal–dielectric contact is assumed to be perfect, and we do not take into account any thermal resistance at the interface. Heat conduction by near-field radiation is neglected, an assumption that is justified for metal particles with sizes of at most a few tens of nanometers.

Coupled Equations

We begin by considering a metal nanoparticle of radius R_p, isolated in a dielectric host medium. This particle absorbs part of the energy of an incident light pulse in a homogeneous way. The equations to be solved describe energy exchanges between the conduction electron gas at temperature $T_\mathrm{e}(t)$, the metal lattice at temperature $T_\mathrm{l}(t)$, and the surrounding matrix at temperature $T_\mathrm{m}(\mathbf{r},t)$ [69,70]. The exciting pulse, the energy source term for the electron gas, is described via $P_\mathrm{abs}(t)$, the instantaneous power absorbed per unit metal volume (given by the characteristics of the pulse, the absorption of the material, and the metal concentration). The first equation then expresses the time dependence of the electron energy due to this source term and the electron–phonon coupling:

$$C_\mathrm{e}\frac{\partial T_\mathrm{e}}{\partial t} = -G(T_\mathrm{e} - T_\mathrm{l}) + P_\mathrm{abs}(t)\ , \tag{6.3}$$

where $C_\mathrm{e} = \gamma_\mathrm{e} T_\mathrm{e}$ is the specific heat of the electron gas, γ_e is a constant depending on the metal, and G is the electron–phonon coupling constant. The second equation

describes the time dependence of the thermal energy of the metal lattice, provided for by the e–ph coupling and transferred to the matrix at the interface:

$$C_{\mathrm{l}}\frac{\partial T_{\mathrm{l}}}{\partial t} = G(T_{\mathrm{e}} - T_{\mathrm{l}}) - \frac{H(t)}{V}\,, \tag{6.4}$$

where C_{l} is the specific heat capacity of the metal lattice, V is the volume of a particle, and $H(t)$ is the instantaneous power transferred to the dielectric matrix from the particle. Let us now consider how to evaluate this last term. It must be handled differently depending on the excitation and observation conditions. One must then identify the heat transport mechanisms from the interface into the matrix. In the general case, at short length and time scales, this transport is described by the Boltzmann equation, or some simplified version of it. This will be the subject of the last section. When the observation time scale (and/or length scale) is long compared with the phonon lifetime τ (respectively, phonon mean free path Λ), it is reasonable to use the Fourier law to describe heat transport. For two of the amorphous dielectrics most commonly used as a matrix in nanocomposites for optics, the values of τ_{ph} and Λ are $\tau_{\mathrm{ph}} \approx 130$ fs and $\Lambda \approx 0.5$ nm for silica (SiO_2) and $\tau_{\mathrm{ph}} \approx 850$ fs and $\Lambda \approx 5.4$ nm for alumina (Al_2O_3).

Classical Diffusive Transport. Fourier Law

In addition to the two equations (6.3) and (6.4), we use the standard parabolic equation for heat diffusion to describe conduction in the dielectric matrix:

$$\frac{\partial T_{\mathrm{m}}}{\partial t} = \frac{\kappa_{\mathrm{m}}}{C_{\mathrm{m}}}\Delta T_{\mathrm{m}}\,, \tag{6.5}$$

where C_{m} and κ_{m} are the specific heat and thermal conductivity of the matrix, respectively. Assuming that $T_{\mathrm{l}}(t) = T_{\mathrm{m}}(R_{\mathrm{p}}, t)$, the function $H(t)$ takes the simple form

$$H(t) = S\kappa_{\mathrm{m}}\left.\frac{\partial T_{\mathrm{m}}}{\partial r}\right|_{r=R_{\mathrm{p}}}\,, \tag{6.6}$$

where S is the surface area of the particle.

Taking the Metal Concentration into Account

This model, originally devised for an isolated particle, has been extended to include the effects of heat exchanges between neighbouring particles within the medium, whilst maintaining the spherical symmetry of the system in order to be able to solve the coupled equations easily, and respecting the values of the metal concentration and average radius of the particles [70]. Apart from the condition already mentioned regarding the observation time relative to the phonon lifetime, the idea of using the

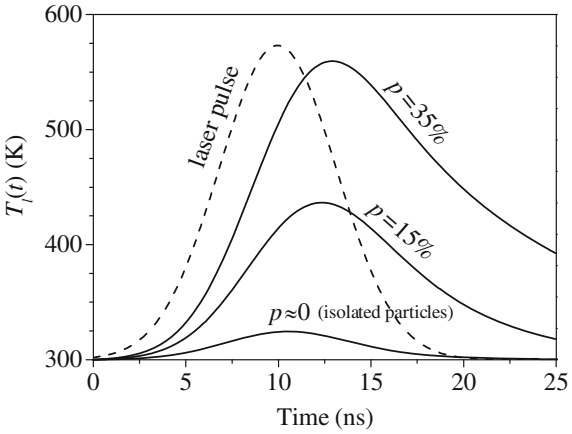

Fig. 6.4 Temperature dynamics of gold nanoparticles ($R = 1.3$ nm) in silica under pulsed laser excitation (pulse width 7.5 ns, peak power absorbed 5×10^{17} W m^{-3}), for three different metal concentrations

Fourier law to describe heat exchanges through the dielectric only remains valid if the distance between neighbouring nanoparticles is large compared with the phonon mean free path.

It has then been shown that, through these exchanges, the metal concentration plays a fundamental role in the temperature dynamics in the slow excitation regime, and also at long times in the ultrafast excitation regime. The main origin of this dependence lies in the fact that, when the heat front emitted by a particle reaches a neighbouring particle, the temperature gradient at the surface of the latter is reduced, which has the effect of slowing down the cooling. This is illustrated in Fig. 6.4 for an Au:SiO_2 nanocomposite in the nanosecond regime.

Ballistic–Diffusive Transport

For very short times and very short distances (and hence for high metal concentrations) or for matrices with higher thermal conductivity, e.g., alumina, dielectric crystals, the Fourier law is no longer applicable. One must then turn to the Boltzmann transport equation (BTE).

The problem of heat transport on small time and length scales has been tackled by several authors [71–77]. An alternative to using the full Boltzmann equation is the so-called ballistic–diffusive equation (BDE), based on the relaxation time approximation and proposed by Chen [73, 76]. It is particularly well suited to the study of transient heat phenomena on the nanoscale, for which it has given similar results to the BTE, but with simpler calculations. The idea is, at each point of the medium, to separate the heat flux $\mathbf{q}(\mathbf{r},t)$, and hence the internal energy $u(\mathbf{r},t)$, into two distinct parts, viz., $\mathbf{q} = \mathbf{q}_\mathrm{b} + \mathbf{q}_\mathrm{d}$ and $u = u_\mathrm{b} + u_\mathrm{d}$. The first represents the contribution of ballistic phonons emitted by the interfaces and the second the contribution of phonons

resulting from scattering processes or reemitted after absorption at other points in the medium.

The main reason for developing this method was to study thin films subjected to a sudden temperature rise on one of their faces [74]. In our case, the BDE can be included in the three-temperature model. To do this, the term in (6.4) describing the energy exchange at the particle–matrix interface is calculated from

$$H(t) = \int_S \mathbf{q}(\mathbf{r},t)\cdot\mathbf{n}\,\mathrm{d}s. \tag{6.7}$$

The flux is integrated over the whole of the particle surface, and $\mathbf{n}$ is the unit vector normal to the surface, directed out into the dielectric.

The phonon intensity in the direction $\boldsymbol{\Omega}$ (analogous to the photon intensity in electromagnetism) is defined by

$$I_\omega(t,\mathbf{r},\boldsymbol{\Omega}) = |\mathbf{v}|\hbar\omega f(t,\mathbf{r},\boldsymbol{\Omega})\frac{\rho(\omega)}{4\pi}\,, \tag{6.8}$$

where $\mathbf{v} = |\mathbf{v}|\boldsymbol{\Omega}$ is the phonon group velocity, ω the phonon angular frequency, f the distribution function, and $\rho(\omega)$ the density of states per unit volume. The heat flux is then found by integrating $I_\omega(t,\mathbf{r},\boldsymbol{\Omega})$ over all phonon energies and in all directions:

$$\mathbf{q}(\mathbf{r},t) = \int_0^\infty \left[\int I_\omega(t,\mathbf{r},\boldsymbol{\Omega})\boldsymbol{\Omega}\,\mathrm{d}\Omega\right]\mathrm{d}\omega\,. \tag{6.9}$$

In the same way, the internal energy is given by

$$u(\mathbf{r},t) = \int_0^\infty \left[\int \frac{I_\omega(t,\mathbf{r},\boldsymbol{\Omega})}{|\mathbf{v}|}\mathrm{d}\Omega\right]\mathrm{d}\omega\,. \tag{6.10}$$

The BTE for the intensity is

$$\frac{\partial I_\omega}{\partial t} + \mathbf{v}\cdot\nabla I_\omega = -\frac{I_\omega - I_{0\omega}}{\tau_\omega}\,, \tag{6.11}$$

where τ_ω is the lifetime of phonons with angular frequency ω, and $I_{0\omega}$ the equilibrium intensity, given as in (6.8) by Bose–Einstein statistics at the equilibrium temperature.

The ballistic component of the flux is related by (6.9) to the ballistic component $I_{\mathrm{b}\omega}$ of the intensity, itself given by the value of I_ω at the point of the interface having emitted in the direction $\boldsymbol{\Omega}$:

$$I_{\mathrm{b}\omega}(t,\mathbf{r},\boldsymbol{\Omega}) = I_\omega\left(t - r'/|\mathbf{v}|, \mathbf{r} - r'\boldsymbol{\Omega}, \boldsymbol{\Omega}\right)\exp\left(-\frac{r'}{\Lambda_\omega}\right)\,, \tag{6.12}$$

where r' is the distance between the point on the interface emitting in the direction $\boldsymbol{\Omega}$ and the point designated by $\mathbf{r}$, and $t - r'/|\mathbf{v}|$ is the delay due to the finite phonon

propagation speed. Λ_ω is of course the mean free path of phonons with angular frequency ω.

Writing down the BTE for $I_{\mathrm{b}\omega}$ [see (6.11)], we obtain the following relation between the ballistic components of the flux and the internal energy:

$$\tau_{\mathrm{ph}} \left(\frac{\partial u_{\mathrm{b}}}{\partial t} + \nabla \cdot \mathbf{q}_{\mathrm{b}} \right) = -u_{\mathrm{b}} \,, \tag{6.13}$$

where τ_{ph} is the average of the relaxation times τ_ω over all frequencies.

To handle the diffusive component, Chen makes an approximation that is commonly used to model thermal radiation, namely, expanding the intensity $I_{\mathrm{d}\omega}$ as a sum of spherical harmonics truncated at order 1 (the P1 method [78]). After integrating over all phonon energies and all directions, the BTE then leads to the relation

$$\tau \frac{\partial \mathbf{q}_{\mathrm{d}}}{\partial t} + \mathbf{q}_{\mathrm{d}} = -\frac{\kappa_{\mathrm{m}}}{C_{\mathrm{m}}} \nabla u_{\mathrm{d}} \,, \tag{6.14}$$

where

$$\kappa_{\mathrm{m}} = \frac{1}{3} \int C_\omega |\mathbf{v}| \Lambda \, \mathrm{d}\omega \approx \frac{1}{3} C_{\mathrm{m}} |\mathbf{v}| \Lambda$$

is the average thermal conductivity, C_ω and C_{m} are the specific heat at angular frequency ω and the total specific heat, respectively, and $\Lambda = |\mathbf{v}| \tau_{\mathrm{ph}}$ is the average of the mean free paths Λ_ω. Using (6.13), taking the divergence of (6.14), and considering energy conservation, viz.,

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{q} = 0 \,, \tag{6.15}$$

with the decomposition of u into its two contributions, $\mathbf{q}_{\mathrm{d}}$ can be eliminated from the equations to arrive finally at

$$\tau_{\mathrm{ph}} \frac{\partial^2 u_{\mathrm{d}}}{\partial t^2} + \frac{\partial u_{\mathrm{d}}}{\partial t} = \nabla \cdot \left(\frac{\kappa_{\mathrm{m}}}{C_{\mathrm{m}}} \nabla u_{\mathrm{d}} \right) - \nabla \cdot \mathbf{q}_{\mathrm{b}} \,. \tag{6.16}$$

This differs from the equation derived in the Cattaneo–Vernotte approach, i.e., the hyperbolic diffusion equation, obtained by adding an inertial term to the Fourier law in order to account for the finite lifetime of the carriers, by the presence of the last term referring to ballistic processes. Chen has shown in the case of thin films that the heat transport as described by this equation is very close to the solution of the Boltzmann equation and very different from the solution given by the Fourier law or the Cattaneo–Vernotte model, revealing the importance of ballistic transport which is not taken into account in the latter two cases [74]. Equations (6.12) and (6.16) are called ballistic–diffusive equations (BDE).

Note that the temperature T_{m} of the matrix does not appear as a variable in the constitutive equations of the BDE, in contrast to the case for standard diffusive transport discussed earlier [see (6.5)]. Indeed, in the ballistic regime, the local carrier distribution is out of equilibrium. The ‘temperature’ does not therefore have its usual

statistical meaning, but can be simply viewed as a measure of the local internal energy [74].

The key feature of the ballistic–diffusive approach is that it assumes that the interfaces only generate ballistic phonons. The interfaces do not contribute to the diffusive component. The only origin for the latter at the boundaries of the medium is thus the flux coming from the medium itself. By virtue of the expansion of $I_{\mathrm{d}\omega}$ as a truncated sum of spherical harmonics and using (6.14), this leads to the following boundary condition for the diffusive component:

$$\tau_{\mathrm{ph}} \frac{\partial u_{\mathrm{d}}}{\partial t} + u_{\mathrm{d}} = \frac{2\Lambda}{3} \nabla u_{\mathrm{d}} \cdot \mathbf{n} \,, \tag{6.17}$$

where $\mathbf{n}$ points out into the dielectric.

In order to minimise the error due to the approximation involved in writing the diffusive component of the intensity, Chen showed that it was preferable to use dimensionless reduced variables, i.e., flux and internal energy differences expressed relative to their equilibrium values, and length and time variables expressed in units of the phonon mean free path and phonon lifetime, respectively.

Application to Core–Shell Particles

Consider a spherical metal nanoparticle subjected to a laser pulse.[7] This particle is surrounded by a shell of dielectric of thickness $d = R_{\mathrm{ex}} - R_{\mathrm{p}}$, where R_{ex} is its outer radius. The whole setup is initially at room temperature T_0. Starting from $t = 0$, the metal particle absorbs part of the energy associated with the incident pulse in the form of electron excitations, energy which then relaxes into the metal lattice. The inner wall of the dielectric shell at $r = R_{\mathrm{p}}$ then emits ballistic phonons at the temperature $T_{\mathrm{l}}(t)$ of the metal lattice.

Applying (6.9) and (6.12), the ballistic flux $q_{\mathrm{b}}(t, \mathbf{r})$ is calculated at a point M of the dielectric shell located at $\mathbf{r}$, at time t, by writing it as the sum of the flux at the equilibrium temperature (the walls emit at temperature T_0), independent of time, and the flux emitted by the particle–shell interface at temperature $\Delta T_{\mathrm{l}} = T_{\mathrm{l}} - T_0$. To do this, we integrate over all points on the two walls (inner and outer) visible from the point M. The same is done for the internal energy of ballistic origin $u_{\mathrm{b}}(t, \mathbf{r})$. Equations (6.7), (6.16), and (6.17) simplify by virtue of the spherical symmetry, since the spatial dependence of the flux and internal energy in the dielectric is radial, and the flux is a radial vector. The joint solution of these equations for heat transport in the dielectric, with ad hoc initial and boundary conditions, and those describing energy exchanges in the metal particle [see (6.3) and (6.4)], determines the dynamics of the electron and metal lattice temperatures, but also the dynamics of the internal energy in the shell [79]. Various boundary conditions can be imposed on the outer surface of the core–shell particle, e.g., thermalisation, adiabaticity, etc.

[7] This work was carried out by our team in collaboration with S. Volz, EM2C, Ecole Centrale Paris, France.

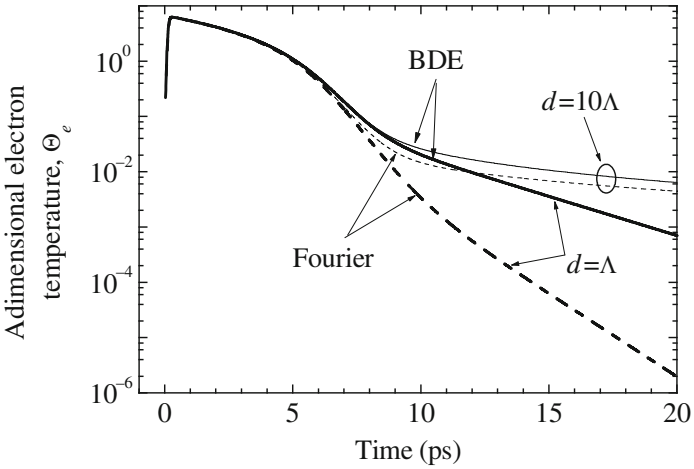

Fig. 6.5 Time dependence of the relative electron temperature, given by the Fourier law (*dashed curves*) and the ballistic–diffusive model (*continuous curves*), for a gold nanoparticle in a dielectric shell of two different thicknesses, illuminated by a 110 fs laser pulse

Figure 6.5 shows the results obtained for a gold particle of radius $R_\mathrm{p} = 10$ nm in an alumina shell (amorphous Al_2O_3) for two thicknesses d, taking total thermalisation as boundary condition on the outer surface (temperature T_0 imposed at $r = R_\mathrm{ex}$). The laser pulse is assumed Gaussian, of duration 110 fs, and centered at $t = 150$ fs. The peak power absorbed per unit volume is 1.4×10^{21} W m^{-3}. The equations were solved by a finite difference method, in which the space and time steps are related in such a way as to ensure stability of the calculation. Room temperature is imposed at the surface of the core–shell particle, which would correspond to the case where it is placed in a medium of high thermal conductivity.

The quantity shown in Fig. 6.5 is the time variation of the reduced electron temperature

$$\Theta_\mathrm{e}(t) = \frac{T_\mathrm{e}(t) - T_0}{T_0} \ ,$$

obtained by the Fourier law and the BDEs, and for two dielectric thicknesses $d = \Lambda$ and $d = 10\Lambda$. We observe to begin with that, at short times $t < 5$ ps, the electron temperature does not depend on the transport mechanism in the dielectric, or indeed the thickness of the latter. In fact, energy relaxation is dominated by electron–phonon coupling within the metal, while the energy transferred to the outside remains negligible. However, after the first few picoseconds, significant differences appear. The relaxation is faster for smaller d, which is not surprising since room temperature is imposed at the outer surface of the dielectric layer. Furthermore, the difference between the two approaches (Fourier and BDE) is greater as the thickness is smaller. For 10Λ, the relaxation of Θ_e is roughly as fast in both cases, reflecting the predominance of the diffusive contribution to heat transport. However, for the thinner shell, the time variation is very different in the two cases. The Fourier law, considering

transport in the dielectric to be diffusive at all times, overestimates the heat exchanges at the interface. Note also that the relaxation gets faster as the thickness decreases, which is due to the higher temperature gradient in the neighbourhood of the metal–dielectric interface, this in turn arising from the thermalisation condition on the outer wall.

The significant differences in behaviour revealed by this example suggest that heat transport in the medium surrounding the particles plays a central role in their cooling dynamics. Among other things, this is reflected in the electron relaxation and hence, as we have seen in Sect. 6.2.3, the optical response of the nanocomposite.

References

1. S.A. Maier, M.L. Brongersma, P.G. Kik, S. Meltzer, A.A.G. Requicha, and H.A. Atwater: Adv. Mater. **13**, 1501–1505 (2001)
2. S. Berciaud, L. Cognet, G.A. Blab, and B. Lounis: Photothermal heterodyne imaging of individual nonfluorescent nanoclusters and nanocrystals, Phys. Rev. Lett. **93**, 257402 (2004)
3. L.R. Hirsch, R.J. Stafford, J.A. Bankson, S.R. Sershen, B. Rivera, R.E. Price, J.D. Hazle, N.J. Halas, and J.L. West: Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance, PNAS **100**, 13549–13554 (2003)
4. C. Loo, A. Lin, L. Hirsch, M. H. Lee, J. Barton, N. Halas, J. West, and R. Drezek: Nanoshell-enabled photonics-based imaging and therapy of cancer, Technol. Cancer Res. Treat. **3**, 33–40 (2004)
5. R.G. Ispasoiu, L. Balogh, O.P. Varnavski, D.A. Tomalia, and T. Goodson III: Large optical limiting from novel metal–dendrimer nanocomposite materials, J. Am. Chem. Soc. **122**, 11005–11006 (2000)
6. L. François, M. Mostafavi, J. Belloni, and J.A. Delaire: Optical limitation induced by gold clusters: Mechanism and efficiency, Phys. Chem. Chem. Phys. **3**, 4965–4971 (2001)
7. S. Papernov and A.W. Schmid: Correlations between embedded single gold nanoparticles in SiO_2 thin film and nanoscale crater formation induced by pulsed-laser radiation, J. Appl. Phys. **92**, 5720–5728 (2002)
8. F. Bonneau, P. Combis, J.L. Rullier, M. Commandré, A. During, J.Y. Natoli, M.J. Pellin, M.R. Savina, E. Cottancin, and M. Pellarin: Observation by photothermal microscopy of increased silica absorption in laser damage induced by gold nanoparticles, Appl. Phys. Lett. **83**, 3855–3857 (2003)
9. B. Bertussi, J.-Y. Natoli, M. Commandré, J.-L. Rullier, F. Bonneau, P. Combis, and P. Bouchut: Photothermal investigation of the laser-induced modification of a single gold nanoparticle in a silica film, Opt. Commun. **254**, 299–309 (2005)
10. H. Ehrenreich and H.R. Philipp: Optical properties of Ag and Cu, Phys. Rev. **128**, 1622–1629 (1962)
11. M.L. Thèye: Investigation of the optical properties of Au by means of thin semi-transparent films, Phys. Rev. B **2**, 3060–3078 (1970)
12. P.B. Johnson and R.W. Christy: Optical constants of the noble metals, Phys. Rev. B **6**, 4370–4379 (1972)
13. D.E. Aspnes, E. Kinsbron, and D.D. Bacon: Optical properties of Au: Sample effects, Phys. Rev. B **21**, 3290–3299 (1980)
14. R.A. Innes and J.R. Sambles: Optical characterisation of gold using surface plasmon-polaritons, J. Phys. F Metal Phys. **17**, 277–287 (1987)
15. D.J. Nash and J.R. Sambles: Surface plasmon-polariton study of the optical dielectric function of silver, J. Mod. Opt. **43**, 81–91 (1996)

16. E.D. Palik (Ed.): *Handbook of Optical Constants of Solids*, Vols. I and II (Academic Press, New York, 1985/1991)
17. N.W. Ashcroft and N.D. Mermin: *Solid State Physics* (Saunders College Pub., 1976)
18. A. Matthiessen: Rep. Brit. Ass. **32**, 144 (1862)
19. C. Voisin: Etude femtoseconde de la dynamique électronique et vibrationnelle dans des nanoparticules métalliques, PhD thesis at Université Paris-Sud (2001)
20. B.R. Cooper, H. Ehrenreich, and H.R. Philipp: Optical properties of noble metals. II, Phys. Rev. **138**, A494–A507 (1965)
21. R. Rosei and D.W. Lynch: Thermomodulation spectra of Al, Au, and Cu, Phys. Rev. B **5**, 3883–3894 (1972)
22. R. Rosei: Temperature modulation of the optical transitions involving the Fermi surface in Ag: Theory, Phys. Rev. B **10**, 474–483 (1974)
23. R. Rosei, C.H. Culp, and J.H. Weaver: Temperature modulation of the optical transitions involving the Fermi surface in Ag: Experimental, Phys. Rev. B **10**, 484–489 (1974)
24. M. Guerrisi, R. Rosei, and P. Winsemius: Splitting of the interband absorption edge in Au, Phys. Rev. B **12**, 557–563 (1975)
25. N.E. Christensen and B.O. Serphin: Relativistic band calculation and the optical properties of gold, Phys. Rev. B **4**, 3321 (1971)
26. G. Mie: Beitrage zur Optik truber Medien, speziell kolloidaler Metallosungen, Ann. Phys. (Leibzig) **25**, 377 (1908)
27. B. Palpant: In: *Nonlinear Optical Properties of Matter: From Molecules to Condensed Phases*, Challenges and Advances in Computational Chemistry and Physics, Vol. 1, pp. 461–508, ed. by M.G. Papadopoulos, J. Leszczynski, and A.J. Sadlej (Springer, 2006)
28. C.F. Bohren and D.P. Huffman: *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983)
29. M. Born and E. Wolf: *Principles of Optics* (Cambridge University Press, Cambridge, 1999)
30. U. Kreibig and M. Vollmer: *Optical Properties of Metal Clusters* (Springer-Verlag, Berlin Heidelberg, 1995)
31. J.C. Maxwell-Garnett: Colours in metal glasses and in metallic films, Philos. Trans. R. Soc. London **203**, 385 (1904)
32. G.P. Pells and M. Shiga: The optical properties of copper and gold as a function of temperature, J. Phys. C **2**, 1835–1846 (1969)
33. R. Rosei, F. Antonangeli, and U.M. Grassano: *d* band positions and widths in gold from very low temperature thermomodulation measurements, Surf. Sci. **37**, 689–699 (1973)
34. P. Winsemius, M. Guerrisi, and R. Rosei: Splitting of the interband absorption edge in Au: Temperature dependence, Phys. Rev. B. **12**, 4570–4572 (1975)
35. P. Winsemius, F.F. van Kampen, H.P. Lengkeek, and C.G. Van Went: Temperature dependence of the optical properties of Au, Ag, and Cu, J. Phys. F Metal Phys. **8**, 1583–1606 (1976)
36. M. Rashidi-Huyeh and B. Palpant: Counterintuitive thermo-optical response of metal–dielectric nanocomposite materials as a result of local electromagnetic field enhancement, Phys. Rev. B **74**, 075405 (2006)
37. B. Palpant, M. Rashidi-Huyeh, B. Gallas, S. Chenot, and S. Fisson: Highly dispersive thermo-optical properties of gold nanoparticles, Appl. Phys. Lett. **90** (22), 223105 (2007)
38. R.H. Magruder, III, R.F. Haglund, Jr., L. Yang, J.E. Wittig, and R.A. Zuhr: Physical and optical properties of Cu nanoclusters fabricated by ion implantation in fused silica, J. Appl. Phys. **76**, 708–715 (1994)
39. K. Puech, W. Blau, A. Grund, C. Bubeck, and G. Gardenas: Picosecond degenerate four-wave mixing in colloidal solutions of gold nanoparticles at high repetition rates, Opt. Lett. **20**, 1613–1615 (1995)
40. H.B. Liao, R.F. Xiao, J.S. Fu, H. Wang, K.S. Wong, and G.K.L. Wong: Origin of third-order optical nonlinearity in Au:SiO_2 composite films on femtosecond and picosecond time scales, Opt. Lett. **23**, 388–390 (1998)
41. M. Falconieri: Thermo-optical effects in Z-scan measurements using high-repetition-rate lasers, J. Opt. A: Pure Appl. Opt. **1**, 662–667 (1999)

42. R.A. Ganeev, A.I. Ryasnyansky, S.R. Kamalov, M.K. Kodirov, and T. Usmanov: Nonlinear susceptibilities, absorption coefficients and refractive indices of colloidal metals, J. Phys. D Appl. Phys. **34**, 1602–1611 (2001)
43. R. de Nalda, R. del Coso, J. Requejo-Isidro, J. Olivares, A. Suarez-Garcia, J. Solis, and C.N. Afonso: Limits to the determination of the nonlinear refractive index by the z-scan method, J. Opt. Soc. Am. B **19**, 289–296 (2002)
44. J.-Y. Bigot, J.-C. Merle, O. Cregut, and A. Daunois: Electron dynamics in copper metallic nanoparticles probed with femtosecond optical pulses, Phys. Rev. Lett. **75**, 4702–4705 (1995)
45. Y. Hamanaka, N. Hayashi, S. Omi, and A. Nakamura: Ultrafast relaxation dynamics of electrons in silver nanocrystals embedded in glass, J. Lumin. **76–77**, 221–225 (1997)
46. M. Nisoli, S. Stagira, S. De Silvestri, A. Stella, P. Tognini, P. Cheyssac, and R. Kofman: Ultrafast electronic dynamics in solid and liquid gallium nanoparticles, Phys. Rev. Lett. **78**, 3575–3578 (1997)
47. M. Perner, P. Bost, U. Becker, M. Mennig, M. Schmitt, and H. Schmidt: Optically induced damping of the surface plasmon resonance in gold colloids, Phys. Rev. Lett. **78**, 2192–2195 (1997); G.V. Hartland, J.H. Hodak, and I. Martini, Comment, ibid. **82**, 3188 (1999); M. Perner, G. von Plessen, and J. Feldmann: Reply, ibid. **82**, 3189 (1999)
48. H. Inouye, K. Tanaka, I. Tanahashi, and K. Hirao: Ultrafast dynamics of nonequilibrium electrons in a gold nanoparticle system, Phys. Rev. B **57**, 11334–11340 (1998)
49. H. Inouye, K. Tanaka, I. Tanahashi, and K. Hirao: Femtosecond optical Kerr effect in the gold nanoparticles system, Jpn. J. Appl. Phys. **37**, L1520–L1522 (1998)
50. S. Link and M.A. El-Sayed: Spectral properties and relaxation dynamics of surface plasmon electronic oscillations in gold and silver nanodots and nanorods, J. Phys. Chem. B **103** (40), 8410–8426 (1999)
51. J.-Y. Bigot, V. Halté, J.-C. Merle, and A. Daunois: Electron dynamics in metallic nanoparticles, Chem. Phys. **251**, 181–203 (2000)
52. N. Del Fatti and F. Vallée: Ultrafast optical nonlinear properties of metal nanoparticles, Appl. Phys. B **73**, 383–390 (2001)
53. Y. Hamanaka, J. Kuwabata, I. Tanahashi, S. Omi, and A. Nakamuka: Ultrafast electron relaxation via breathing vibration of gold nanocrystals embedded in a dielectric medium, Phys. Rev. B **63**, 104302 (2001)
54. C. Voisin, N. Del Fatti, D. Christofilos, and F. Vallée: Ultrafast electron dynamics and optical nonlinearities in metal nanoparticles, J. Phys. Chem. B **105**, 2264–2280 (2001)
55. W.S. Fann, R. Storz, H.W.K. Tom, and J. Bokor: Electron thermalization in gold, Phys. Rev. B **46**, 13592–13595 (1992)
56. C.-K. Sun, F. Vallée, L. Acioli, E.P. Ippen, and J.G. Fujimoto: Femtosecond investigation of electron thermalization in gold, Phys. Rev. B **48**, 12365–12368 (1993); Femtosecond-tunable measurement of electron thermalization in gold, ibid. **50**, 15337–15348 (1994)
57. R.H.M. Groeneveld, R. Sprik, and A. Lagendijk: Femtosecond spectroscopy of electron–electron and electron–phonon energy relaxation in Ag and Au, Phys. Rev. B **51**, 11433–11445 (1995)
58. N. Del Fatti, R. Bouffanais, F. Vallée, and C. Flytzanis: Nonequilibrium electron interactions in metal films, Phys. Rev. Lett. **81**, 922–925 (1998)
59. G. Tas and H.J. Maris: Electron diffusion in metals studied by picosecond ultrasonics, Phys. Rev. B **49**, 15046–15054 (1994)
60. C. Suárez, W.E. Bron, and T. Juhasz: Dynamics and transport of electronic carriers in thin gold films, Phys. Rev. Lett. **75**, 4536–4539 (1995)
61. V.E. Gusev and O.B. Wright: Ultrafast nonequilibrium dynamics of electrons in metals, Phys. Rev. B **57**, 2878–2888 (1998)
62. Y. Guillet: Dynamique de la réponse optique ultrarapide d'une assemblée de nanoparticules d'or, PhD thesis at Université Pierre et Marie Curie, Paris 6 (2007)
63. D. Pines and P. Nozières: *The Theory of Quantum Liquids*, Vol. I: *Normal Fermi Liquids* (W.A. Benjamin Inc., New York, 1966)

64. J. Lermé, G. Celep, M. Broyer, E. Cottancin, M. Pellarin, A. Arbouet, D. Christofilos, C. Guillon, P. Langot, N. Del Fatti, and F. Vallée: Effects of confinement on the electron and lattice dynamics in metal nanoparticles, Eur. Phys. J. D **34**, 199–204 (2005)
65. R.H. Ritchie: Coupled electron–hole cascade in a free electron gas, J. Appl. Phys. **37**, 2276–2278 (1966)
66. C. Voisin, D. Christofilos, N. Del Fatti, F. Vallée, B. Prével, E. Cottancin, J. Lermé, M. Pellarin, and M. Broyer: Size-dependent electron–electron interactions in metal nanoparticles, Phys. Rev. Lett. **85**, 2200–2203 (2000)
67. A. Arbouet, C. Voisin, D. Christofilos, P. Langot, N. Del Fatti, F. Vallée, J. Lermé, G. Celep, E. Cottancin, M. Gaudry, M. Pellarin, M. Broyer, M. Maillard, M.-P. Pileni, and M. Treguer: Electron–phonon scattering in metal clusters, Phys. Rev. Lett. **90**, 177401 (2003)
68. E.D. Belotkii, S.N. Luk'yanets, and P.M. Tomchuk: Theory of hot electrons in island metal films, Sov. Phys. JETP **74**, 88–94, 1992
69. Y. Hamanaka, J. Kuwabata, I. Tanahashi, S. Omi, and A. Nakamuka: Ultrafast electron relaxation via breathing vibration of gold nanocrystals embedded in a dielectric medium, Phys. Rev. B **63**, 104302 (2001)
70. M. Rashidi-Huyeh and B. Palpant: Thermal response of nanocomposite materials under pulsed laser excitation, J. Appl. Phys. **96**, 4475 (2004)
71. A.A. Joshi and A. Majumdar: Transient ballistic and diffusive phonon heat transport in thin films, J. Appl. Phys. **74**, 31–39 (1993)
72. G. Chen: Nonlocal and nonequilibrium heat conduction in the vicinity of nanoparticles, J. Heat Transfer **118**, 539–545 (1996)
73. G. Chen: Ballistic–diffusive heat-conduction equations, Phys. Rev. Lett. **86**, 2297–2300 (2001)
74. G. Chen: Ballistic–diffusive equations for transient heat conduction from nano- to macroscales, J. Heat Transfer ASME **124**, 320–328 (2002)
75. D.G. Cahill, W.K. Ford, K.E. Goodson, G.D. Mahan, A. Majumdar, H.J. Maris, R. Merlin, and S.R. Phillpot: Nanoscale thermal transport, J. Appl. Phys. **93**, 793–818 (2003)
76. G. Chen, D. Borca-Tasciuc, and R. Yang: Nanoscale heat transfer. In: *Encyclopedia of Nanoscience and Nanotechnology* Vol. X, pp. 1–30, ed. by H.S. Nalwa (Am. Sci. Publ., 2004)
77. S. Volz (Ed.): *Microscale and Nanoscale Heat Transfer*, Topics in Applied Physics, Vol. 107 (Springer, 2007)
78. D. Lemonnier: Solution of the Boltzmann equation for phonon transport. In [77], pp. 77–106
79. M. Rashidi-Huyeh, S. Volz, and B. Palpant: Non-Fourier heat transport in metal–dielectric core–shell nanoparticles under ultrafast laser pulse excitation, Phys. Rev. B **78**, 125408 (2008)

Chapter 7
Mie Theory and the Discrete Dipole Approximation. Calculating Radiative Properties of Particulate Media, with Application to Nanostructured Materials

Franck Enguehard

7.1 Introduction

Radiative transfer in a semi-transparent medium can be described by a spacetime dependent directional monochromatic specific intensity field $L_\lambda(\mathbf{r},\mathbf{n},t)$, where λ is the wavelength, $\mathbf{r}$ the field point, $\mathbf{n}$ the unit direction vector, and t the time. This field $L_\lambda(\mathbf{r},\mathbf{n},t)$ obeys an integro-differential equation called the radiative transfer equation (RTE) which has the general form [1]:

$$\frac{1}{c_\lambda}\frac{\partial L_\lambda(\mathbf{r},\mathbf{n},t)}{\partial t}+\mathbf{n}\cdot\nabla_{\mathbf{r}}L_\lambda(\mathbf{r},\mathbf{n},t)=-(\kappa_\lambda+\sigma_\lambda)L_\lambda(\mathbf{r},\mathbf{n},t)+\kappa_\lambda n_\lambda^2 L_\lambda^0\big(T(\mathbf{r},t)\big)+\frac{\sigma_\lambda}{4\pi}\int_{4\pi}\Phi_\lambda(\mathbf{n}',\mathbf{n})L_\lambda(\mathbf{r},\mathbf{n}',t)\mathrm{d}\Omega' . \quad (7.1)$$

In this formulation, c_λ is the speed of energy propagation in the semi-transparent medium, while $\nabla_{\mathbf{r}}$ is the gradient with respect to position $\mathbf{r}$, n_λ is the refractive index, i.e., the real part of the complex optical index m_λ of the medium, $T(\mathbf{r},t)$ is the temperature field in the medium, and $L_\lambda^0(T)$ is the specific intensity of the equilibrium radiation at temperature T. Finally, κ_λ, σ_λ, and $\Phi_\lambda(\mathbf{n}',\mathbf{n})$ are the bulk radiative properties of the medium, viz., its absorption coefficient, scattering coefficient, and scattering phase function, respectively. Introducing the extinction coefficient $\beta_\lambda=\kappa_\lambda+\sigma_\lambda$ and scattering albedo $\omega_\lambda=\sigma_\lambda/\beta_\lambda$, the steady-state version of the RTE (7.1) (valid on time scales such that the propagation of radiation can be assumed instantaneous) can be written in the form

$$\frac{1}{\beta_\lambda}\mathbf{n}\cdot\nabla_{\mathbf{r}}L_\lambda(\mathbf{r},\mathbf{n},t)=-L_\lambda(\mathbf{r},\mathbf{n},t)+(1-\omega_\lambda)n_\lambda^2 L_\lambda^0\big(T(\mathbf{r},t)\big)+\frac{\omega_\lambda}{4\pi}\int_{4\pi}\Phi_\lambda(\mathbf{n}',\mathbf{n})L_\lambda(\mathbf{r},\mathbf{n}',t)\mathrm{d}\Omega' . \quad (7.2)$$

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 151
DOI 10.1007/978-3-642-04258-4_7, © Springer-Verlag Berlin Heidelberg 2009

Apart from the field $T(\mathbf{r},t)$, the input data for this equation reduce to the radiative properties β_λ, ω_λ, and $\Phi(\mathbf{n}',\mathbf{n})$ of the medium. Furthermore, since the radiative flux vector $\phi^{\mathrm{R}}(\mathbf{r},t)$ is calculated from spectral and directional integration of the vectorial intensity field $L_\lambda(\mathbf{r},\mathbf{n},t)\mathbf{n}$, it follows that evaluation of the spatiotemporal field $\phi^{\mathrm{R}}(\mathbf{r},t)$ requires prior knowledge of the radiative properties β_λ, ω_λ, and $\Phi(\mathbf{n}',\mathbf{n})$ of the semi-transparent medium over all wavelengths λ of the spectral band relevant to the temperature in the medium [2].

In this chapter, we will be concerned with semi-transparent media that are not homogeneous but particulate, i.e., made up of a solid or liquid dispersed in the form of particles throughout a host medium that does not absorb radiation. Some examples of this family of heterogeneous semi-transparent media are:

- clouds, ensembles of water droplets or ice crystals suspended in air,
- smoke, tiny solid particles suspended in air,
- the atmosphere, a population of gas molecules behaving like very small particles,
- nanostructured materials made up of agglomerated solid nanoparticles.

We shall see that there are theoretical tools for calculating the spectra of the radiative properties β_λ, ω_λ, and $\Phi(\mathbf{n}',\mathbf{n})$ of such media. In the following, we shall outline two of these techniques based on Mie theory and the discrete dipole approximation.

The chapter is organised into three main sections. To begin with, we describe the theoretical aspects of the interaction between radiation and a particle or population of particles. We shall present the results of this theory applicable to the general case of a particle with arbitrary shape. We then focus on the special case of a spherical particle. The main formulas of Mie theory will be given, and the results of this theory illustrated by calculating the radiative properties of a cloud of water droplets over a broad spectrum of wavelengths. Finally, we shall discuss the discrete dipole approximation and show that this modelling technique can to some extent fill an important gap in Mie theory, namely, the fact that the latter cannot account for the spatial arrangement of matter in the radiative properties it produces.

Before beginning, we would like to draw the reader's attention to the following two points:

- All the theoretical developments to follow will be monochromatic, i.e., they will concern a single given wavelength λ. To simplify notation, the index λ reminding us of the monochromatic nature of the various quantities will thus be omitted.
- All the electromagnetic fields to be manipulated here will be harmonic in time. In accordance with the convention to be found in most standard textbooks on optics, we shall associate with these fields a time dependence going as $\exp(-\mathrm{i}\omega t)$, rather than as $\exp(\mathrm{i}\omega t)$, in which $\mathrm{i} = \sqrt{-1}$, and ω and t are the angular frequency and time, respectively. This is an important point, because it requires the complex optical index m to have a positive imaginary part. This means that, throughout this chapter, the complex optical index m will be expressed in the form $m = n + \mathrm{i}\chi$, in which n and χ are the refractive index and extinction index, respectively. It should be noted that analytic expressions of Mie theory based on the convention $\exp(\mathrm{i}\omega t)$ are sometimes found in the literature, whence the complex optical index takes the form $m = n - \mathrm{i}\chi$, and the resulting formulas differ slightly from those presented here.

7.2 Absorption and Scattering by a Particle of Arbitrary Shape and by a Population of Such Particles

Most of the results discussed in this section have been adapted from Chap. 3 of *Absorption and Scattering of Light by Small Particles*, by C.F. Bohren and D.R. Huffman [3], which is without doubt a key bibliographical reference in this field.

7.2.1 Incident Electromagnetic Field, Poynting Vector, and Associated Power

To begin with, consider a particle (labelled 1) with specified shape, size, and constitution, isolated in a 3D space filled with some non-absorbent host medium (labelled 2). The particle is illuminated by an incident electromagnetic wave propagating in medium 2, and which can, without loss of generality, be assumed to be a plane wave and monochromatic, with wavelength λ_2, angular frequency ω, and wave number $k_2 = 2\pi/\lambda_2 = \omega/c_2$. The electromagnetic field $\{\mathbf{E}_{\text{inc}}(\mathbf{r},t), \mathbf{H}_{\text{inc}}(\mathbf{r},t)\}$ associated with the incident wave can thus be written in the form

$$\begin{cases} \mathbf{E}_{\text{inc}}(\mathbf{r},t) = E_0 \mathbf{e} \exp\left[\mathrm{i}(k_2 \mathbf{u}\cdot\mathbf{r} - \omega t)\right] , \\ \nabla \times \mathbf{E}_{\text{inc}} = -\dfrac{\partial \mathbf{B}_{\text{inc}}}{\partial t} , \quad \mathbf{H}_{\text{inc}} = \dfrac{\mathbf{B}_{\text{inc}}}{\mu_2} \quad \Longrightarrow \quad \mathbf{H}_{\text{inc}}(\mathbf{r},t) = \dfrac{k_2}{\mu_2 \omega}\left[\mathbf{u} \times \mathbf{E}_{\text{inc}}(\mathbf{r},t)\right] . \end{cases} \tag{7.3}$$

In these relations, μ_2 is the magnetic permeability of medium 2, while the wavelength λ_2 and speed of light c_2 in this medium are given by $\lambda_2 = \lambda_0/n_2$ and $c_2 = c_0/n_2$, with λ_0 and c_0 the corresponding properties in vacuum, and n_2 the refractive index of medium 2 at the wavelength λ_0, whence the wave number k_2 can be written either $k_2 = 2\pi n_2/\lambda_0$ or $k_2 = \omega n_2/c_0$. Further, **E**, **B**, and **H** are the electric field, magnetic induction, and magnetic field, respectively. The quantity E_0 is the amplitude in $\mathrm{V\,m^{-1}}$ of the incident electric field $\mathbf{E}_{\text{inc}}$. This quantity is taken to be real, without loss of generality, since choosing a complex value would merely shift the phase term $k_2\mathbf{u}\cdot\mathbf{r} - \omega t$ by a constant. The unit vector **e** gives the polarisation of the field $\mathbf{E}_{\text{inc}}$. In contrast with the amplitude E_0, the polarisation can sometimes assume complex values, as happens when the incident wave is not linearly but elliptically polarised. In this case, the condition that **e** should be a unit vector is written $\|\mathbf{e}\| = 1$, extending the definition of the norm to complex vectors in the following way:

$$\|\mathbf{X}\| = \sqrt{\mathbf{X}\cdot\mathbf{X}^*} , \tag{7.4}$$

where $\mathbf{X}^*$ is the complex conjugate of **X**. Finally, the vector **u**, also of unit length ($\|\mathbf{u}\| = 1$), gives the direction of propagation of the incident wave. Unlike the polarisation vector **e**, this vector **u** is always real. In addition, **u** is orthogonal to **e**, expressed by $\mathbf{u}\cdot\mathbf{e} = 0$.

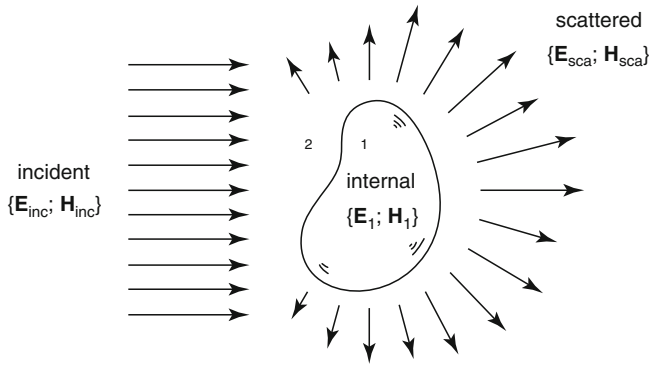

Fig. 7.1 Incident, internal (within the particle), and scattered electromagnetic fields. Taken from [3]

A power per unit area $\psi(\mathbf{r},t)$ can be associated with the electromagnetic field $\{\mathbf{E}_{\text{inc}}(\mathbf{r},t),\mathbf{H}_{\text{inc}}(\mathbf{r},t)\}$ given in (7.3). In the most general case, with arbitrary electromagnetic field $\{\mathbf{E}(\mathbf{r},t),\mathbf{H}(\mathbf{r},t)\}$, this power per unit area is calculated from the Poynting vector $\mathbf{S}(\mathbf{r},t) = \mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)$ of this field. The vector $\mathbf{S}$ defined in this way, which has physical dimensions $\mathrm{W\,m^{-2}}$, specifies the magnitude and direction of the electromagnetic energy flux at any point and time. When the field $\{\mathbf{E}(\mathbf{r},t),\mathbf{H}(\mathbf{r},t)\}$ varies sinusoidally with time, i.e., as $\exp(-\mathrm{i}\omega t)$, it can be shown that the Poynting vector $\mathbf{S}(\mathbf{r},t)$ also varies sinusoidally about an average value $\langle\mathbf{S}(\mathbf{r})\rangle$ given by

$$\langle\mathbf{S}(\mathbf{r})\rangle = \frac{1}{2}\mathrm{Re}\left[\mathbf{E}(\mathbf{r},t)\times\mathbf{H}^{*}(\mathbf{r},t)\right] . \tag{7.5}$$

This is the only quantity that most measurement devices are able to evaluate. Applying this result (7.5) to the incident electromagnetic field $\{\mathbf{E}_{\text{inc}}(\mathbf{r},t),\mathbf{H}_{\text{inc}}(\mathbf{r},t)\}$ given in (7.3), we obtain

$$\langle\mathbf{S}_{\text{inc}}(\mathbf{r})\rangle = \frac{k_2 E_0^2}{2\mu_2\omega}\mathbf{u} . \tag{7.6}$$

The time-averaged Poynting vector $\langle\mathbf{S}_{\text{inc}}(\mathbf{r})\rangle$ is thus a constant vector, with norm $k_2E_0^2/2\mu_2\omega$, everywhere oriented along the direction of propagation $\mathbf{u}$. The power per unit area ψ_{inc} associated with the incident illumination is therefore given by

$$\boxed{\psi_{\text{inc}} = \frac{k_2 E_0^2}{2\mu_2\omega}} \tag{7.7}$$

7.2.2 Electromagnetic Fields Within and Scattered by the Particle

Under the effect of the incident field $\{\mathbf{E}_{\text{inc}}(\mathbf{r},t),\mathbf{H}_{\text{inc}}(\mathbf{r},t)\}$, an electromagnetic field $\{\mathbf{E}_1(\mathbf{r},t),\mathbf{H}_1(\mathbf{r},t)\}$ will be generated within the particle. In addition, the

interaction between the particle and the incident radiation will produce an electromagnetic field scattered by the particle into medium 2 (see Fig. 7.1). This field, denoted by $\{\mathbf{E}_{\rm sca}(\mathbf{r},t),\mathbf{H}_{\rm sca}(\mathbf{r},t)\}$, will be superposed upon the incident field $\{\mathbf{E}_{\rm inc}(\mathbf{r},t),\mathbf{H}_{\rm inc}(\mathbf{r},t)\}$ in medium 2, in such a way that the total electromagnetic field within medium 2 will be the field

$$\left\{\mathbf{E}_2(\mathbf{r},t)=\mathbf{E}_{\rm inc}(\mathbf{r},t)+\mathbf{E}_{\rm sca}(\mathbf{r},t),\mathbf{H}_2(\mathbf{r},t)=\mathbf{H}_{\rm inc}(\mathbf{r},t)+\mathbf{H}_{\rm sca}(\mathbf{r},t)\right\}.$$

The problem at this point is to determine the two unknown fields, viz., the field $\{\mathbf{E}_1(\mathbf{r},t),\mathbf{H}_1(\mathbf{r},t)\}$ in the particle (medium 1) and the field $\{\mathbf{E}_{\rm sca}(\mathbf{r},t),\mathbf{H}_{\rm sca}(\mathbf{r},t)\}$ in the host medium (medium 2). These two fields each satisfy the four Maxwell equations without the free charge and current terms. Since the time dependence of our fields is sinusoidal, i.e., going as $\exp(-{\rm i}\omega t)$, these four equations are:

$$\boxed{\nabla\cdot\mathbf{E}_\alpha=0\,,\quad \nabla\cdot\mathbf{H}_\alpha=0\,,\quad \nabla\times\mathbf{E}_\alpha={\rm i}\omega\mu_\beta\mathbf{H}_\alpha\,,\quad \nabla\times\mathbf{H}_\alpha=-{\rm i}\omega\varepsilon_\beta\mathbf{E}_\alpha} \tag{7.8}$$

where ε is the dielectric permittivity and the pair (α,β) is equal to (1,1) for the field $\{\mathbf{E}_1(\mathbf{r},t),\mathbf{H}_1(\mathbf{r},t)\}$ and (sca,2) for the field $\{\mathbf{E}_{\rm sca}(\mathbf{r},t),\mathbf{H}_{\rm sca}(\mathbf{r},t)\}$. The set of equations (7.8) comes with continuity conditions at the interface between the particle and the host medium. Naturally, the latter concern the total fields in the two media, i.e., the field $\{\mathbf{E}_1(\mathbf{r},t),\mathbf{H}_1(\mathbf{r},t)\}$ in the particle and the field $\{\mathbf{E}_2(\mathbf{r},t),\mathbf{H}_2(\mathbf{r},t)\}$ in the surrounding medium. They express the continuity of the tangential components of the electric and magnetic fields on either side of the interface. Let $\mathbf{R}$ be the position vector of an arbitrary point on the interface and $\mathbf{N}$ the unit normal vector to the interface at $\mathbf{R}$, oriented toward the outside of the particle. Then the continuity conditions at the interface can be written in the form

$$\boxed{\left[\mathbf{E}_2(\mathbf{R},t)-\mathbf{E}_1(\mathbf{R},t)\right]\times\mathbf{N}=\mathbf{0}\,,\qquad \left[\mathbf{H}_2(\mathbf{R},t)-\mathbf{H}_1(\mathbf{R},t)\right]\times\mathbf{N}=\mathbf{0}} \tag{7.9}$$

Equations (7.8) and (7.9) constitute a complete mathematical system. Hence, the problem of determining the two unknown electromagnetic fields $\{\mathbf{E}_1(\mathbf{r},t),\mathbf{H}_1(\mathbf{r},t)\}$ and $\{\mathbf{E}_{\rm sca}(\mathbf{r},t),\mathbf{H}_{\rm sca}(\mathbf{r},t)\}$ is formally solved. Unfortunately, when the shape of the particle has no particular geometrical symmetry, it is impossible to find analytical solutions for these two fields, and one must resort to numerical techniques.

7.2.3 Extinction, Absorbed, and Scattered Power

We assume at this point that the field $\{\mathbf{E}_{\rm sca}(\mathbf{r},t),\mathbf{H}_{\rm sca}(\mathbf{r},t)\}$ scattered by the particle into the host medium is known. We now construct a fictional sphere Σ with large enough radius to completely enclose the particle. The electromagnetic power Ψ_2 crossing this sphere Σ toward the inside is given by minus the flux of the Poynting vector $\langle\mathbf{S}_2(\mathbf{r})\rangle$ in medium 2 through Σ:

$$\Psi_2 = -\int_\Sigma \langle \mathbf{S}_2(\mathbf{R})\rangle \cdot \mathbf{N} \mathrm{d}\Sigma \,, \tag{7.10}$$

where the position vector $\mathbf{R}$ is an arbitrary point on the sphere and the vector $\mathbf{N}$ is the unit vector normal to the sphere at $\mathbf{R}$ and pointing toward the outside of the sphere. Since the Poynting vector $\langle \mathbf{S}_2(\mathbf{r})\rangle$ is given by

$$\langle \mathbf{S}_2(\mathbf{r})\rangle = \frac{1}{2}\mathrm{Re}\left[\mathbf{E}_2(\mathbf{r},t)\times\mathbf{H}_2^*(\mathbf{r},t)\right] \,,$$

and the electromagnetic field $\{\mathbf{E}_2(\mathbf{r},t),\mathbf{H}_2(\mathbf{r},t)\}$ results from the superposition of the incident and scattered fields, it follows that the power Ψ_2 decomposes into three terms as follows:

$$\Psi_2 = \Psi_{\mathrm{inc}} - \Psi_{\mathrm{sca}} + \Psi_{\mathrm{coupling}} \,, \tag{7.11}$$

where

$$\Psi_{\mathrm{inc}} = -\int_\Sigma \langle \mathbf{S}_{\mathrm{inc}}(\mathbf{R})\rangle \cdot \mathbf{N} \mathrm{d}\Sigma = -\frac{1}{2}\int_\Sigma \mathrm{Re}\left[\mathbf{E}_{\mathrm{inc}}(\mathbf{R},t)\times\mathbf{H}_{\mathrm{inc}}^*(\mathbf{R},t)\right]\cdot\mathbf{N}\mathrm{d}\Sigma \,,$$

$$\Psi_{\mathrm{sca}} = +\int_\Sigma \langle \mathbf{S}_{\mathrm{sca}}(\mathbf{R})\rangle \cdot \mathbf{N} \mathrm{d}\Sigma = +\frac{1}{2}\int_\Sigma \mathrm{Re}\left[\mathbf{E}_{\mathrm{sca}}(\mathbf{R},t)\times\mathbf{H}_{\mathrm{sca}}^*(\mathbf{R},t)\right]\cdot\mathbf{N}\mathrm{d}\Sigma \,, \tag{7.12}$$

$$\Psi_{\mathrm{coupling}} = -\frac{1}{2}\int_\Sigma \mathrm{Re}\Big\{\left[\mathbf{E}_{\mathrm{inc}}(\mathbf{R},t)\times\mathbf{H}_{\mathrm{sca}}^*(\mathbf{R},t)\right] + \left[\mathbf{E}_{\mathrm{sca}}(\mathbf{R},t)\times\mathbf{H}_{\mathrm{inc}}^*(\mathbf{R},t)\right]\Big\}\cdot\mathbf{N}\mathrm{d}\Sigma \,.$$

Up to a sign, the term Ψ_{inc} is the flux of the incident Poynting vector through the sphere Σ. Since this sphere lies entirely within the host medium, which is assumed to be non-absorbent, it follows that the power Ψ_{inc} can only be zero. The equality (7.11) thus reduces to

$$\Psi_{\mathrm{coupling}} = \Psi_2 + \Psi_{\mathrm{sca}} \,, \tag{7.13}$$

in which the two powers Ψ_2 and Ψ_{sca} on the right-hand side are both positive or zero:

- In terms of sign, the power Ψ_2 is defined as being supplied to the interior of Σ. Ψ_2 is the flux toward the interior of Σ of the Poynting vector $\langle \mathbf{S}_2(\mathbf{r})\rangle$ associated with the total electromagnetic field $\{\mathbf{E}_2(\mathbf{r},t),\mathbf{H}_2(\mathbf{r},t)\}$ within the host medium, and this flux can only be positive or zero. Indeed, $\Psi_2 < 0$ would mean that globally the power associated with the total field were leaving the closed surface Σ, which would only be possible if the material volume enclosed within Σ were able to produce electromagnetic energy, a situation we do not consider here. Consequently, $\Psi_2 \geq 0$ represents the power absorbed by the volume bounded by the fictional sphere Σ. Since this volume comprises only the particle and some of the surrounding medium, which has been assumed to be non-absorbent, it follows that Ψ_2, always non-negative, is the power absorbed by the particle. We shall thus denote this power Ψ_2 by Ψ_{abs} from this point.
- The power Ψ_{sca} given by (7.12) has a very simple physical interpretation. It is the signed flux toward the outside of Σ of the Poynting vector $\langle \mathbf{S}_{\mathrm{sca}}(\mathbf{r})\rangle$ associated with the scattered electromagnetic field $\{\mathbf{E}_{\mathrm{sca}}(\mathbf{r},t),\mathbf{H}_{\mathrm{sca}}(\mathbf{r},t)\}$. This power Ψ_{sca} is thus simply the power of the electromagnetic radiation scattered by the particle.

To end this section, let us consider the power Ψ_{coupling} as given in (7.12). The Poynting vector

$$\left\langle \mathbf{S}_{\text{coupling}}(\mathbf{r}) \right\rangle = \frac{1}{2}\text{Re}\left\{ \left[\mathbf{E}_{\text{inc}}(\mathbf{r},t) \times \mathbf{H}^{*}_{\text{sca}}(\mathbf{r},t) \right] + \left[\mathbf{E}_{\text{sca}}(\mathbf{r},t) \times \mathbf{H}^{*}_{\text{inc}}(\mathbf{r},t) \right] \right\}$$

associated with this power is remarkable in that it couples the incident and scattered electromagnetic fields. Ψ_{coupling} can thus be interpreted as the power resulting from interactions between the two electromagnetic fields that coexist in the host medium. Furthermore, the power Ψ_{coupling} is positive since it is equal to the sum of the powers Ψ_{abs} and Ψ_{sca}, which are themselves both positive. It thus represents power supplied to the interior of Σ, or a power that disappears from the exterior of Σ.

So to sum up the main results described here, we have two important points to bear in mind:

- Since the host medium in which the particle resides has been assumed to be non-absorbent, it follows that the power Ψ_{coupling} is equal to the sum of the two powers Ψ_{abs} and Ψ_{sca}.
- $\Psi_{\text{abs}} \geq 0$ is the power absorbed by the particle, while $\Psi_{\text{sca}} \geq 0$ is the power scattered by the particle. The power $\Psi_{\text{coupling}} \geq 0$ is commonly referred to as the extinction power of the particle. It will henceforth be denoted Ψ_{ext}. Here, the word 'extinction' implies the disappearance of a photon from its initial trajectory, either by absorption or by deviation, where deviation is synonymous with scattering.

7.2.4 Expressing the Extinction and Scattered Powers in Terms of the Incident and Scattered Electric Fields

The expression for the power Ψ_{ext} in terms of the incident and scattered electromagnetic fields can be found from (7.12), viz., the formula for Ψ_{coupling}. This expression can be rewritten entirely in terms of the electric fields $\mathbf{E}_{\text{inc}}(\mathbf{r},t)$ and $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$ by appealing to a very useful mathematical result, a special case of a theorem that has been around for over a hundred years, namely, the optical theorem. This result is quoted without proof in [3], and the authors stress that it is obtained after a considerable amount of algebra. We shall therefore simply accept it here. It then turns out that the expression in (7.12) for the power Ψ_{ext} can be rewritten in the form

$$\Psi_{\text{ext}} = \frac{2\pi\varepsilon_2 c_2}{k_2^2}\text{Re}\left\{ \left[\left(\frac{\mathbf{E}_{\text{inc}}(\mathbf{r},t)}{\exp\left[\text{i}(k_2\mathbf{u}\cdot\mathbf{r}-\omega t)\right]} \right)^{*} \cdot \left(\frac{\mathbf{E}_{\text{sca}}(\mathbf{r},t)}{\exp\left[\text{i}(k_2 r-\omega t)\right]/(-\text{i}k_2 r)} \right) \right]_{\substack{r\to+\infty \\ \theta=0}} \right\} \tag{7.14}$$

This new formula for Ψ_{ext} requires some explanations:

- The term in square brackets in (7.14) must be evaluated for $r \to \infty$ and $\theta = 0$, where r is the norm of the position vector $\mathbf{r}$, i.e., $r = \left\|\mathbf{r}\right\|$, and θ is the angle between $\mathbf{r}$ and the direction of propagation $\mathbf{u}$ of the incident illumination ($0 \leq \theta \leq \pi$). As a consequence, (7.14) involves the properties of the electric field scattered (i) into the so-called radiative or far-field region, reached whenever the product $k_2 r$ becomes much bigger than unity, and (ii) into a rather particular scattering direction, namely straight forward, without deviation.
- The term

$$\frac{\mathbf{E}_{\text{inc}}(\mathbf{r},t)}{\exp\left[\mathrm{i}(k_2\mathbf{u}\cdot\mathbf{r}-\omega t)\right]}$$

appearing in (7.14) represents the incident electric field without its spacetime dependence. To some extent, the same goes for the term

$$\frac{\mathbf{E}_{\text{sca}}(\mathbf{r},t)}{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]/(-\mathrm{i}k_2 r)}$$

when the scattered electric field $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$ is considered in the far field. Indeed, in the radiative region, this field $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$ is approximately transverse, i.e., $\mathbf{E}_{\text{sca}}(\mathbf{r},t)\cdot\mathbf{r} \approx 0$, and has an asymptotic form of the type [4]

$$\boxed{\mathbf{E}_{\text{sca}}(\mathbf{r},t) \sim \frac{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]}{-\mathrm{i}k_2 r}\mathbf{A}_{\text{sca}}(\mathbf{n})} \tag{7.15}$$

in which the vector $\mathbf{n}$ is the unit vector parallel to and in the same direction as $\mathbf{r}$, i.e., $\mathbf{n} = \mathbf{r}/r$, and the vector field $\mathbf{A}_{\text{sca}}(\mathbf{n})$, the amplitude vector field of the scattered electric field $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$, is no longer a function of the position vector $\mathbf{r}$, but only its direction $\mathbf{n}$. The far-field term

$$\frac{\mathbf{E}_{\text{sca}}(\mathbf{r},t)}{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]/(-\mathrm{i}k_2 r)}$$

appearing in (7.14) thus arises as the scattered electric field with its distance (r) and time (t) dependence removed, but keeping its dependence on the direction $\mathbf{n}$, because it is none other than the amplitude vector field $\mathbf{A}_{\text{sca}}(\mathbf{n})$.

Let us return for a moment to a point brought up earlier. Equation (7.14) for the extinction power Ψ_{ext} involves the properties of the scattered electric field $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$ in a single and rather special scattering direction, namely straight forward, without deviation, corresponding to the zero value of the scattering angle θ, and to a scattering direction $\mathbf{n}$ equal to the direction of propagation $\mathbf{u}$. This last point is a rather curious consequence of the optical theorem. Indeed, the extinction power appears to depend only on the forward scattering amplitude, whereas physically, extinction is conceived of as the combined effect of absorption by the particle and scattering by the particle in all space directions! An interpretation of this curious consequence of the optical theorem is suggested in [3].

Let us now consider the power $\Psi_{\rm sca}$ scattered by the particle, as given by the second relation of (7.12). It is quite legitimate to let the radius of the fictional integration sphere Σ tend to infinity in this formula. The expression (7.15) for the field $\mathbf{E}_{\rm sca}(\mathbf{r},t)$ in the far field is then applicable. Concerning the scattered magnetic field $\mathbf{H}_{\rm sca}(\mathbf{r},t)$ in the far field, the expression for this field follows from the Maxwell equation $\nabla\times\mathbf{E}_{\rm sca}=\mathrm{i}\omega\mu_2\mathbf{H}_{\rm sca}$ and the specific features of the electromagnetic field $\{\mathbf{E}_{\rm sca}(\mathbf{r},t),\mathbf{H}_{\rm sca}(\mathbf{r},t)\}$ in the radiative zone:

$$\nabla\times\mathbf{E}_{\rm sca}=\mathrm{i}\omega\mu_2\mathbf{H}_{\rm sca}\;,\quad \nabla\times\mathbf{E}_{\rm sca}=\mathrm{i}k_2\mathbf{n}\times\mathbf{E}_{\rm sca}\quad\Longrightarrow\quad \mathbf{H}_{\rm sca}=\frac{k_2}{\omega\mu_2}\mathbf{n}\times\mathbf{E}_{\rm sca}\;. \tag{7.16}$$

The Poynting vector $\langle\mathbf{S}_{\rm sca}(\mathbf{r})\rangle$ can then be found immediately. Recalling that in the far field $\mathbf{E}_{\rm sca}(\mathbf{r},t)\cdot\mathbf{n}=0$, this leads to

$$\langle\mathbf{S}_{\rm sca}(\mathbf{r})\rangle=\frac{k_2}{2\omega\mu_2}\big\|\mathbf{E}_{\rm sca}(\mathbf{r},t)\big\|^2\mathbf{n}=\frac{1}{2\omega\mu_2k_2r^2}\big\|\mathbf{A}_{\rm sca}(\mathbf{n})\big\|^2\mathbf{n}\;, \tag{7.17}$$

whence the power $\Psi_{\rm sca}$ is given by

$$\Psi_{\rm sca}=\frac{1}{2\omega\mu_2k_2R^2}\int_\Sigma\big\|\mathbf{A}_{\rm sca}(\mathbf{N})\big\|^2\mathrm{d}\Sigma\;. \tag{7.18}$$

The above surface integral can be transformed into a directional integral over 4π steradians. Indeed, since $\mathrm{d}\Sigma$ is a surface element of the fictional sphere Σ of radius R, the quantity $\mathrm{d}\Sigma/R^2$ is the elementary solid angle $\mathrm{d}\Omega$, and in the end we obtain the following expression for the power $\Psi_{\rm sca}$:

$$\boxed{\Psi_{\rm sca}=\frac{1}{2\omega\mu_2k_2}\int_{4\pi}\big\|\mathbf{A}_{\rm sca}(\mathbf{n})\big\|^2\mathrm{d}\Omega} \tag{7.19}$$

7.2.5 *Extinction, Absorption, and Scattering Cross-Sections. Associated Efficiencies and Scattering Phase Function*

The extinction cross-section $C_{\rm ext}$ of the particle is defined as the ratio of the power $\Psi_{\rm ext}$ extinguished by the particle (as defined above) and the power per unit area $\psi_{\rm inc}$ associated with the incident illumination, given in (7.7). The quantity $C_{\rm ext}$ thus has dimensions of area. The absorption and scattering cross-sections of the particle, denoted by $C_{\rm abs}$ and $C_{\rm sca}$, respectively, are defined in a completely analogous way, viz., $C_{\rm abs}=\Psi_{\rm abs}/\psi_{\rm inc}$ and $C_{\rm sca}=\Psi_{\rm sca}/\psi_{\rm inc}$, respectively, so that the relation $\Psi_{\rm ext}=\Psi_{\rm abs}+\Psi_{\rm sca}$ between the three powers implies the relation $C_{\rm ext}=C_{\rm abs}+C_{\rm sca}$ between the three cross-sections.

When the particle is spherical with radius a, the cross-sections are often expressed relative to its geometrical cross-section πa^2. This leads one to introduce new,

dimensionless quantities called efficiencies. The extinction efficiency Q_{ext} is thus defined as $C_{\text{ext}}/\pi a^2$, and likewise for the absorption and scattering efficiencies, viz., $Q_{\text{abs}} = C_{\text{abs}}/\pi a^2$ and $Q_{\text{sca}} = C_{\text{sca}}/\pi a^2$, respectively. Once again, the three efficiencies obviously satisfy the relation $Q_{\text{ext}} = Q_{\text{abs}} + Q_{\text{sca}}$.

The expression for the scattering phase function of the particle follows directly from the expression (7.19) for the scattered power Ψ_{sca}. Indeed, as this expression (7.19) is written in the form of a directional integral over 4π steradians, and as the scattering phase function $\Phi(\mathbf{n})$ is defined as the direction indicator of the probability of scattering in the direction $\mathbf{n}$, it follows that this function $\Phi(\mathbf{n})$ is simply equal, up to a multiplicative constant, to the function $\left\|\mathbf{A}_{\text{sca}}(\mathbf{n})\right\|^2$ appearing under the integral in the expression (7.19) for the scattered power Ψ_{sca}. Moreover, if this function $\Phi(\mathbf{n})$ is normalised by the condition $\int_{4\pi} \Phi(\mathbf{n}) \mathrm{d}\Omega = 4\pi$ usually applied in radiative transfer calculations, then the expression for the scattering phase function of the particle is found immediately to be

$$\boxed{\Phi(\mathbf{n}) = \frac{2\pi}{\omega \mu_2 k_2 \Psi_{\text{sca}}} \left\|\mathbf{A}_{\text{sca}}(\mathbf{n})\right\|^2} \tag{7.20}$$

7.2.6 Directions of Propagation and Polarisation

At this point, it should be recalled that the theory we have just developed and the results produced from it are based on the assumption of an incident illumination $\{\mathbf{E}_{\text{inc}}(\mathbf{r},t), \mathbf{H}_{\text{inc}}(\mathbf{r},t)\}$ propagating in a given direction $\mathbf{u}$ with given polarisation $\mathbf{e}$ [see (7.3) for the form of this incident illumination]. As a consequence, the various quantities we have introduced so far, namely, the powers Ψ_{ext}, Ψ_{abs}, and Ψ_{sca}, the cross-sections C_{ext}, C_{abs}, and C_{sca}, and the efficiencies Q_{ext}, Q_{abs}, and Q_{sca}, together with the scattering phase function $\Phi(\mathbf{n})$, all depend on the two vectors $\mathbf{u}$ and $\mathbf{e}$. We shall thus denote them by $\Psi_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{e})$, and so on, with a superscript P to remind us that the incident electromagnetic field is polarised. Now thermal radiation is not polarised. Our next task is therefore to extend the definition of the ten quantities listed above to this type of radiation.

Regarding the powers, cross-sections, and efficiencies, the non-polarised nature of the incident illumination is taken into account as follows. Let us subject our particle to an incident electromagnetic field $\{\mathbf{E}_{\text{inc}}(\mathbf{r},t), \mathbf{H}_{\text{inc}}(\mathbf{r},t)\}$ with specified direction of propagation $\mathbf{u}$ but unpolarised, or more exactly, with polarisation $\mathbf{e}$ fluctuating randomly and equiprobably with regard to its direction in the plane orthogonal to $\mathbf{u}$. Let $\Psi_{\text{ext}}^{\text{NP}}(\mathbf{u})$ be the particle extinction power in this illumination configuration. In this notation $\Psi_{\text{ext}}^{\text{NP}}(\mathbf{u})$, the superscript NP reminds us that the incident electromagnetic field is not polarised, and the dependence of the extinction power on the direction of propagation $\mathbf{u}$ is indicated explicitly. It can be shown (using arguments similar to those found for example in [5] to derive the expression for the reflection coefficient in the presence of non-polarised incident illumination) that this extinction power $\Psi_{\text{ext}}^{\text{NP}}(\mathbf{u})$ is equal to the arithmetic mean of the two powers $\Psi_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{v})$

and $\Psi^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{w})$ that the particle would have 'extinguished' if it had been subjected successively to two polarised incident illuminations with the same direction of propagation $\mathbf{u}$ and power per unit area ψ_{inc}, with polarisations $\mathbf{v}$ and $\mathbf{w}$ naturally orthogonal to $\mathbf{u}$, but also orthogonal to one another:

$$\boxed{\Psi^{\mathrm{NP}}_{\mathrm{ext}}(\mathbf{u}) = \frac{\Psi^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{v}) + \Psi^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{w})}{2}} \tag{7.21}$$

Note in passing that it can also be shown that the sum

$$\Psi^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{v}) + \Psi^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{w})$$

is independent of the choice of the pair of vectors $(\mathbf{v},\mathbf{w})$ making up an orthonormal basis with the propagation vector $\mathbf{u}$. Dividing (7.21) by the incident power per unit area ψ_{inc}, we obtain the extinction cross-section of the particle in the presence of unpolarised radiation:

$$C^{\mathrm{NP}}_{\mathrm{ext}}(\mathbf{u}) = \frac{\Psi^{\mathrm{NP}}_{\mathrm{ext}}(\mathbf{u})}{\psi_{\mathrm{inc}}} = \frac{C^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{v}) + C^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{w})}{2}\,. \tag{7.22}$$

The results (7.21) and (7.22) for the extinction power and cross-section extend to the absorption and scattering powers and cross-sections, in such a way that the relations

$$\Psi^{\mathrm{NP}}_{\mathrm{ext}}(\mathbf{u}) = \Psi^{\mathrm{NP}}_{\mathrm{abs}}(\mathbf{u}) + \Psi^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u})\,, \qquad C^{\mathrm{NP}}_{\mathrm{ext}}(\mathbf{u}) = C^{\mathrm{NP}}_{\mathrm{abs}}(\mathbf{u}) + C^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u})$$

remain valid. An expression for the scattering phase function $\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n})$ under non-polarised illumination is found from the expression for the associated scattered power $\Psi^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u})$, viz.,

$$\Psi^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u}) = \frac{\Psi^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{v}) + \Psi^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{w})}{2}\,,$$

which implies

$$\boxed{\Psi^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u}) = \frac{1}{4\omega\mu_2 k_2}\int_{4\pi}\left[\left\|\mathbf{A}^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{v},\mathbf{n})\right\|^2 + \left\|\mathbf{A}^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{w},\mathbf{n})\right\|^2\right]\mathrm{d}\Omega} \tag{7.23}$$

with the obvious notation, whence

$$\boxed{\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n}) = \frac{\pi}{\omega\mu_2 k_2 \Psi^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u})}\left[\left\|\mathbf{A}^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{v},\mathbf{n})\right\|^2 + \left\|\mathbf{A}^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{w},\mathbf{n})\right\|^2\right]} \tag{7.24}$$

7.2.7 Radiative Properties of a Population of Particles

To end the general considerations of Sect. 7.2, we shall discuss the interaction between some incident radiation and a whole population of discrete particles, rather than just one, all assumed to have the same shape and size, and to be made up of the same material and oriented in the same way. We also assume that the particles are distributed randomly within the host medium with uniform probability density. Let n be the number of particles per unit volume.

The extinction coefficient β of the cloud of particles is by definition the cumulative extinction cross-section (under non-polarised light) per unit volume. When defined in this way, β has dimensions m^{-1}. If we assume that each particle in the cloud interacts with the incident radiation exactly as though it were isolated within the host medium (this is the so-called independent matter–radiation interaction regime), then it can be shown immediately that the extinction coefficient $\beta(\mathbf{u})$, which is in principle a function of the direction of illumination $\mathbf{u}$, can be expressed very simply in terms of the extinction cross-section $C_{\mathrm{ext}}^{\mathrm{NP}}(\mathbf{u})$ of a single particle under non-polarised illumination via the relation

$$\boxed{\beta(\mathbf{u}) = nC_{\mathrm{ext}}^{\mathrm{NP}}(\mathbf{u})} \tag{7.25}$$

This result extends to the absorption and scattering coefficients $\kappa(\mathbf{u})$ and $\sigma(\mathbf{u})$ of the particle cloud (quantities that are in principle directional), i.e., $\kappa(\mathbf{u}) = nC_{\mathrm{abs}}^{\mathrm{NP}}(\mathbf{u})$ and $\sigma(\mathbf{u}) = nC_{\mathrm{sca}}^{\mathrm{NP}}(\mathbf{u})$, respectively, so that:

- We have the relation $\beta(\mathbf{u}) = \kappa(\mathbf{u}) + \sigma(\mathbf{u})$.
- The directional albedo $\omega(\mathbf{u})$ is given by

$$\omega(\mathbf{u}) = \frac{\sigma(\mathbf{u})}{\beta(\mathbf{u})} = \frac{C_{\mathrm{sca}}^{\mathrm{NP}}(\mathbf{u})}{C_{\mathrm{ext}}^{\mathrm{NP}}(\mathbf{u})} \,.$$

Finally, since the particles are assumed to be identical and oriented in the same way throughout the cloud, they all have the same scattering phase function $\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n})$. This function is therefore also the scattering phase function for the whole particle cloud [6].

The independent matter–radiation interaction regime is clearly the easiest to understand physically. It corresponds to the situation where the average distance between any two particles in the cloud is large enough for the interaction between the electromagnetic field and a given particle to occur without any memory of interactions the field may have had previously with other particles. It is also clear that, if the particle density is high enough, the independence hypothesis is likely to fail, and the relations of the type given in (7.25) will no longer hold. We shall return to this question in the following section devoted to Mie theory.

7.3 Mie Theory

The name of Mie theory refers to the analytic solution of the problem of electromagnetism described in Sect. 7.2.2, i.e., the determination of the electromagnetic fields $\{\mathbf{E}_1(\mathbf{r},t),\mathbf{H}_1(\mathbf{r},t)\}$ inside the particle and $\{\mathbf{E}_2(\mathbf{r},t),\mathbf{H}_2(\mathbf{r},t)\}$ outside the particle, when the particle is spherical. This theory is named after the German physicist Gustav Mie, who published these calculations in 1908 [7] as a result of his studies of the absorption and scattering properties of aqueous suspensions of gold colloidal particles in the visible light range.

From the analytic solution of Mie's electromagnetic problem, we will be able to deduce expressions for the extinction and scattering cross-sections, as well as the scattering phase function. The transition from the radiative response of an isolated particle to that of a whole population of particles will then be discussed. This will once again raise the question of dependence effects in the matter–radiation interaction, already mentioned in Sect. 7.2.7. The difficulties related to a non-uniform distribution of diameters within the particle population will also be addressed. Finally, Sect. 7.3 ends by exemplifying a typical Mie calculation to evaluate the radiative response of a particle cloud.

As we shall soon see, the mathematical expressions of Mie theory are somewhat involved. In order to simplify the notation, we have thus decided to express all analytic expressions in Sect. 7.3 without explicit inclusion of the temporal harmonic term $\exp(-\mathrm{i}\omega t)$, which will be understood.

7.3.1 Analytic Solution to Mie's Electromagnetic Problem

The route to this solution involves some rather tedious mathematics. Good accounts can be found in [3, 8], and the appendix at the end of this chapter provides a summary of the four main steps in the solution. For the present purposes, we shall simply quote the analytic expression for the scattered electromagnetic field $\{\mathbf{E}_{\mathrm{sca}}(\mathbf{r},t),\mathbf{H}_{\mathrm{sca}}(\mathbf{r},t)\}$. This result suffices to express some of the quantities relevant to the particle–radiation interaction introduced in Sect 7.2, namely, the extinction and scattering powers and cross-sections, together with the scattering phase function.

The spherical particle of radius a is assumed to be illuminated by a polarised incident wave, with direction of propagation $\mathbf{u}$ and polarisation $\mathbf{e}$, where $\mathbf{e} \perp \mathbf{u}$. The incident electric field $\mathbf{E}_{\mathrm{inc}}(\mathbf{r})$ is then given by

$$\mathbf{E}_{\mathrm{inc}}(\mathbf{r}) = E_0 \mathbf{e} \exp(\mathrm{i}k_2 \mathbf{u}\cdot\mathbf{r}) \,. \tag{7.26}$$

The geometrical symmetry of the particle naturally suggests adopting a spherical coordinate system (r,θ,φ), where the angle θ $(0 \le \theta \le \pi)$ is defined by $\theta = (\mathbf{u},\mathbf{r})$ and the angle φ is the polar angle between $\mathbf{e}$ and the orthogonal projection of $\mathbf{r}$ in the plane orthogonal to $\mathbf{u}$ $(0 \le \varphi \le 2\pi)$. In this coordinate system, the scattered electromagnetic field $\{\mathbf{E}_{\mathrm{sca}}(r,\theta,\varphi),\mathbf{H}_{\mathrm{sca}}(r,\theta,\varphi)\}$ resulting from the Mie theory has the following analytic form:

$$\boxed{\begin{cases} \mathbf{E}_{\text{sca}}(r,\theta,\varphi) = \displaystyle\sum_{n=1}^{\infty} E_n \Big[-b_n \mathbf{M}^{h2}_{\text{o1}n}(r,\theta,\varphi) + \mathrm{i} a_n \mathbf{N}^{h2}_{\text{e1}n}(r,\theta,\varphi) \Big] , \\ \mathbf{H}_{\text{sca}}(r,\theta,\varphi) = \dfrac{k_2}{\mathrm{i}\omega\mu_2} \displaystyle\sum_{n=1}^{\infty} E_n \Big[-b_n \mathbf{N}^{h2}_{\text{o1}n}(r,\theta,\varphi) + \mathrm{i} a_n \mathbf{M}^{h2}_{\text{e1}n}(r,\theta,\varphi) \Big] , \end{cases}} \tag{7.27}$$

with

$$\mathbf{H}_{\text{sca}}(r,\theta,\varphi) = \frac{\nabla \times \mathbf{E}_{\text{sca}}(r,\theta,\varphi)}{\mathrm{i}\omega\mu_2} .$$

The notation here requires some explanation:

- The coefficients E_n ($n \geq 1$) are given by

$$E_n = \mathrm{i}^n E_0 \frac{2n+1}{n(n+1)} .$$

- The two sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, usually called the Mie sequences, have rather complicated expressions which simplify somewhat in the particular case where the particle and the host medium have the same magnetic permeability. The terms a_n and b_n are then calculated using the following relations:

$$\boxed{\begin{aligned} a_n &= \frac{m\psi_n(mx)\psi_n'(x) - \psi_n(x)\psi_n'(mx)}{m\psi_n(mx)\xi_n'(x) - \xi_n(x)\psi_n'(mx)} , \\ b_n &= \frac{\psi_n(mx)\psi_n'(x) - m\psi_n(x)\psi_n'(mx)}{\psi_n(mx)\xi_n'(x) - m\xi_n(x)\psi_n'(mx)} , \end{aligned}} \tag{7.28}$$

  where the functions $\psi_n(\rho)$ and $\xi_n(\rho)$ are the Ricatti–Bessel functions defined by

$$\psi_n(\rho) = \rho\, j_n(\rho) , \qquad \xi_n(\rho) = \rho\, h_n(\rho) , \tag{7.29}$$

  with j_n and h_n the spherical Bessel and Hankel functions (see the Appendix for more details). The quantity x, defined by $x = k_2 a$, is a dimensionless parameter commonly called the size parameter relative to the host medium (medium 2). It is real because the host medium is assumed non-absorbent. The quantity m, defined by $m = m_1/m_2 = k_1/k_2$, is another dimensionless parameter, representing the contrast of the complex optical index between the particle and the host medium. The product mx is equal to $k_1 a$, the size parameter relative to the particle (medium 1). It may be complex, if the particle is made from a material that absorbs radiation. A glance at (7.28) shows that the two sequences of terms a_n and b_n are functions only of the two dimensionless parameters introduced here, viz., the size parameter x (real) and the contrast m of the complex optical indices (possibly complex).
- Finally, the functions $\mathbf{M}^{h2}_{\text{e1}n}(r,\theta,\varphi)$, $\mathbf{M}^{h2}_{\text{o1}n}(r,\theta,\varphi)$, $\mathbf{N}^{h2}_{\text{e1}n}(r,\theta,\varphi)$, and $\mathbf{N}^{h2}_{\text{o1}n}(r,\theta,\varphi)$ are called vector spherical harmonics (see the appendix for more details), with the following analytic expressions:

$$\mathbf{M}^{h2}_{\mathrm{e1}n}(r,\theta,\varphi) = -h_n(k_2 r)\frac{P_n^1(\cos\theta)}{\sin\theta}\sin\varphi\,\mathbf{e}_\theta - h_n(k_2 r)\frac{\mathrm{d}P_n^1(\cos\theta)}{\mathrm{d}\theta}\cos\varphi\,\mathbf{e}_\varphi \tag{7.30}$$

$$\mathbf{M}^{h2}_{\mathrm{o1}n}(r,\theta,\varphi) = h_n(k_2 r)\frac{P_n^1(\cos\theta)}{\sin\theta}\cos\varphi\,\mathbf{e}_\theta - h_n(k_2 r)\frac{\mathrm{d}P_n^1(\cos\theta)}{\mathrm{d}\theta}\sin\varphi\,\mathbf{e}_\varphi \tag{7.31}$$

$$\begin{aligned}\mathbf{N}^{h2}_{\mathrm{e1}n}(r,\theta,\varphi) = {} & n(n+1)\frac{h_n(k_2 r)}{k_2 r}P_n^1(\cos\theta)\cos\varphi\,\mathbf{e}_r \\ & + \frac{1}{k_2 r}\frac{\mathrm{d}\left[r h_n(k_2 r)\right]}{\mathrm{d}r}\frac{\mathrm{d}P_n^1(\cos\theta)}{\mathrm{d}\theta}\cos\varphi\,\mathbf{e}_\theta \\ & - \frac{1}{k_2 r}\frac{\mathrm{d}\left[r h_n(k_2 r)\right]}{\mathrm{d}r}\frac{P_n^1(\cos\theta)}{\sin\theta}\sin\varphi\,\mathbf{e}_\varphi\end{aligned} \tag{7.32}$$

$$\begin{aligned}\mathbf{N}^{h2}_{\mathrm{o1}n}(r,\theta,\varphi) = {} & n(n+1)\frac{h_n(k_2 r)}{k_2 r}P_n^1(\cos\theta)\sin\varphi\,\mathbf{e}_r \\ & + \frac{1}{k_2 r}\frac{\mathrm{d}\left[r h_n(k_2 r)\right]}{\mathrm{d}r}\frac{\mathrm{d}P_n^1(\cos\theta)}{\mathrm{d}\theta}\sin\varphi\,\mathbf{e}_\theta \\ & + \frac{1}{k_2 r}\frac{\mathrm{d}\left[r h_n(k_2 r)\right]}{\mathrm{d}r}\frac{P_n^1(\cos\theta)}{\sin\theta}\cos\varphi\,\mathbf{e}_\varphi\end{aligned} \tag{7.33}$$

in which P_n^1 are the associated Legendre functions of the first kind and $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\varphi)$ is the local right-handed orthonormal basis associated with the spherical coordinates (r,θ,φ).

7.3.2 Extinction and Scattering Cross-Sections. Scattering Phase Function

Given the analytic expression (7.27) for the field $\{\mathbf{E}_{\mathrm{sca}}(r,\theta,\varphi), \mathbf{H}_{\mathrm{sca}}(r,\theta,\varphi)\}$ scattered by the spherical particle, we may use the results (7.14) and (7.19) in Sect. 7.2.4 to calculate the powers $\Psi^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{e})$ and $\Psi^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{e})$ extinguished and scattered by the particle.

The expression for the field $\mathbf{E}_{\mathrm{sca}}(r,\theta,\varphi)$ in the far field is obtained from (7.27) using the following results:

$$h_n(\rho) \sim \frac{(-\mathrm{i})^n \exp(\mathrm{i}\rho)}{\mathrm{i}\rho} \quad \text{and} \quad h_n'(\rho) \sim \frac{(-\mathrm{i})^n \exp(\mathrm{i}\rho)}{\rho}\,, \quad \text{as } \rho \to \infty\,. \tag{7.34}$$

The calculation eventually yields

$$\mathbf{E}_{\text{sca}}(r,\theta,\varphi) \sim E_0 \frac{\exp(\mathrm{i}k_2 r)}{-\mathrm{i}k_2 r}\Big[S_2(\cos\theta)\cos\varphi\,\mathbf{e}_\theta - S_1(\cos\theta)\sin\varphi\,\mathbf{e}_\varphi\Big]\,, \quad \text{as } r\longrightarrow\infty\,, \tag{7.35}$$

where S_1 and S_2 are defined by

$$\begin{cases} S_1(\cos\theta) = \displaystyle\sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}\Big[a_n\pi_n(\cos\theta) + b_n\tau_n(\cos\theta)\Big]\,, \\ S_2(\cos\theta) = \displaystyle\sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}\Big[a_n\tau_n(\cos\theta) + b_n\pi_n(\cos\theta)\Big]\,, \end{cases} \tag{7.36}$$

and π_n and τ_n are themselves defined by

$$\pi_n(\cos\theta) = \frac{P_n^1(\cos\theta)}{\sin\theta}\,, \qquad \tau_n(\cos\theta) = \frac{\mathrm{d}P_n^1(\cos\theta)}{\mathrm{d}\theta}\,. \tag{7.37}$$

At this point, the extinction power $\Psi_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{e})$ can be calculated immediately from (7.14). Given that

$$\pi_n(1) = \tau_n(1) = \frac{n(n+1)}{2}\,,$$

the result is

$$\boxed{\Psi_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{e}) = \frac{\pi\varepsilon_2 c_2 E_0^2}{k_2^2}\sum_{n=1}^{\infty}(2n+1)\text{Re}(a_n+b_n)} \tag{7.38}$$

whence it transpires that $\Psi_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{e})$ has no dependence on either the direction of propagation $\mathbf{u}$ or the polarisation $\mathbf{e}$. This is precisely what one would expect, given the geometrical symmetry of the spherical particle. Equation (7.38) is thus also the expression for the power $\Psi_{\text{ext}}^{\text{NP}}$ extinguished by the particle under non-polarised illumination, whatever the direction of incidence of this illumination, so that, dividing $\Psi_{\text{ext}}^{\text{NP}}$ first by the incident power per unit area ψ_{inc} as given in (7.7), then by the geometrical cross-section πa^2 of the particle, we obtain first the expression for the extinction cross-section $C_{\text{ext}}^{\text{NP}}$ of the particle, and then its associated efficiency $Q_{\text{ext}}^{\text{NP}}$:

$$\boxed{C_{\text{ext}}^{\text{NP}} = \frac{2\pi}{k_2^2}\sum_{n=1}^{\infty}(2n+1)\text{Re}(a_n+b_n)\,, \quad Q_{\text{ext}}^{\text{NP}} = \frac{2}{x^2}\sum_{n=1}^{\infty}(2n+1)\text{Re}(a_n+b_n)} \tag{7.39}$$

The expression for the power $\Psi_{\text{sca}}^{\text{P}}(\mathbf{u},\mathbf{e})$ scattered by the particle follows directly from (7.19) in Sect. 7.2.4. According to (7.35), the amplitude vector field $\mathbf{A}_{\text{sca}}^{\text{P}}(\mathbf{n})$ defined in (7.15) is given by

$$\mathbf{A}_{\text{sca}}^{\text{P}}(\theta,\varphi) = E_0\Big[S_2(\cos\theta)\cos\varphi\,\mathbf{e}_\theta - S_1(\cos\theta)\sin\varphi\,\mathbf{e}_\varphi\Big]\,. \tag{7.40}$$

The power $\Psi_{\text{sca}}^{\text{P}}(\mathbf{u},\mathbf{e})$ is thus given by

$$\Psi_{\text{sca}}^{\text{P}}(\mathbf{u},\mathbf{e}) = \frac{E_0^2}{2\omega\mu_2 k_2}\int_{4\pi}\Big[\big|S_2(\cos\theta)\big|^2\cos^2\varphi + \big|S_1(\cos\theta)\big|^2\sin^2\varphi\Big]\mathrm{d}\Omega\,, \tag{7.41}$$

which depends neither on $\mathbf{u}$ nor on $\mathbf{e}$, but only on the size parameter x and the complex optical index contrast m. This power is thus also equal to the power $\Psi_{\text{sca}}^{\text{NP}}$ scattered by the particle under non-polarised illumination, whatever the direction of incidence of the illumination. Continuing the calculation of the integral (7.41) and using the special orthogonality properties of the functions π_n and τ_n, the expression for $\Psi_{\text{sca}}^{\text{NP}}$ reduces to the following simplified formula:

$$\Psi_{\text{sca}}^{\text{NP}} = \frac{\pi \varepsilon_2 c_2 E_0^2}{k_2^2} \sum_{n=1}^{\infty} (2n+1)\left(|a_n|^2 + |b_n|^2\right) , \tag{7.42}$$

whence the expressions for the scattering cross-section and scattering efficiency of the particle can be written in the form

$$\boxed{C_{\text{sca}}^{\text{NP}} = \frac{2\pi}{k_2^2} \sum_{n=1}^{\infty} (2n+1)\left(|a_n|^2 + |b_n|^2\right) , \quad Q_{\text{sca}}^{\text{NP}} = \frac{2}{x^2} \sum_{n=1}^{\infty} (2n+1)\left(|a_n|^2 + |b_n|^2\right)} \tag{7.43}$$

Finally, consider the scattering phase function $\Phi^{\text{NP}}(\mathbf{u},\mathbf{n})$ under non-polarised illumination, which has the general formula (7.24) in Sect. 7.2.6. The particular symmetry properties of a spherical particle allow one to transform this general expression (7.24) to the form

$$\Phi^{\text{NP}}(\mathbf{u},\mathbf{n}) = \frac{\pi}{\omega \mu_2 k_2 \Psi_{\text{sca}}^{\text{NP}}} \left[\left\| \mathbf{A}_{\text{sca}}^{\text{P}}(\theta,\varphi) \right\|^2 + \left\| \mathbf{A}_{\text{sca}}^{\text{P}}(\theta,\varphi+\pi/2) \right\|^2 \right] , \tag{7.44}$$

where the amplitude vector field $\mathbf{A}_{\text{sca}}^{\text{P}}(\theta,\varphi)$ is given in (7.40). Straightforward calculation leads finally to

$$\boxed{\Phi^{\text{NP}}(\mathbf{u},\mathbf{n}) = \frac{|S_1(\cos\theta)|^2 + |S_2(\cos\theta)|^2}{\sum_{n=1}^{\infty}(2n+1)\left(|a_n|^2 + |b_n|^2\right)} = \frac{2}{x^2 Q_{\text{sca}}^{\text{NP}}} \left[|S_1(\cos\theta)|^2 + |S_2(\cos\theta)|^2 \right]} \tag{7.45}$$

and it transpires that the scattering phase function $\Phi^{\text{NP}}(\mathbf{u},\mathbf{n})$ expressed in this way has cylindrical symmetry, i.e., it does not depend on the angle φ, but only on the angle θ between the scattering direction $\mathbf{n}$ and the incidence direction $\mathbf{u}$.

7.3.3 A Special Case: Rayleigh Scattering

Several limiting cases of the general Mie theory are often described in the literature [3, 6, 8]. One of these is Rayleigh scattering, when x and $|m-1|x$ are both much smaller than unity, or Rayleigh–Gans scattering, when $|m-1|$ and $|m-1|x$ are both much smaller than unity, or again the case of geometrical optics (ray tracing) and the theory of diffraction when $x \gg 1$ (the list is not exhaustive). In his textbook [8], H.C. Van de Hulst provides a map of the (m,x) plane (m being the optical index and x the size parameter), showing the positions of the various special cases for the interaction

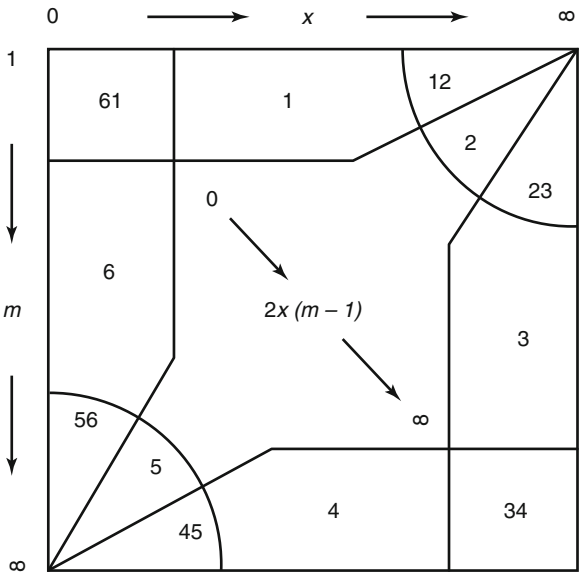

Fig. 7.2 Map of the (m,x) plane, where m is the optical index and x the size parameter, showing different special cases for the interaction between electromagnetic radiation and a non-absorbing spherical particle (hence with real optical index m). Taken from [8]

Table 7.1 Table accompanying Fig. 7.2. Adapted from [8]

Region	x	$m-1$	$x(m-1)$	Special case	Extinction formula
61	Small	Small	Small		$Q = 32(m-1)^2x^4/27$
1	Arbitrary	Small	Small	Rayleigh–Gans	
12	Large	Small	Small		$Q = 2(m-1)^2x^2$
2	Large	Small	Arbitrary	Anomalous diffraction	
23	Large	Small	Large		$Q = 2$
3	Large	Arbitrary	Large	Large spheres	
34	Large	Large	Large		$Q = 2$
4	Arbitrary	Large	Large	Total reflector	
45	Small	Large	Large		$Q = 10x^4/3$
5	Small	Large	Arbitrary	Optical resonance	
56	Small	Large	Small		$Q = 8x^4/3$
6	Small	Arbitrary	Small	Rayleigh scattering	

between electromagnetic radiation and a non-absorbing spherical particle (hence with real optical index m). This map is shown in Fig. 7.2 (see also Table 7.1).

We shall examine Rayleigh scattering in some detail. This phenomenon explains why the sky looks blue, and why the sun turns red as it sets. In these examples, visible light is scattered by gas molecules in the atmosphere. When x and $|m-1|x$ are both much smaller than 1, the same is also true of $|m|x$. It can then be shown that the scattering coefficient a_1 can be approximated by

$$a_1 \sim -\frac{2\mathrm{i}x^3}{3}\frac{m^2-1}{m^2+2}\,, \tag{7.46}$$

and that all the other scattering coefficients a_n ($n \geq 2$) and b_n ($n \geq 1$) are infinitely small compared with a_1. It follows that (7.39), (7.43), and (7.45) for the extinction efficiency, scattering efficiency, and scattering phase function then reduce to

$$\boxed{Q_{\mathrm{ext}}^{\mathrm{NP}} = \frac{6}{x^2}\mathrm{Re}(a_1) = 4x\mathrm{Im}\left(\frac{m^2-1}{m^2+2}\right)\,, \quad Q_{\mathrm{sca}}^{\mathrm{NP}} = \frac{6}{x^2}|a_1|^2 = \frac{8}{3}x^4\left|\frac{m^2-1}{m^2+2}\right|^2} \tag{7.47}$$

$$\boxed{\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n}) = \frac{2}{x^2 Q_{\mathrm{sca}}^{\mathrm{NP}}}\frac{9}{4}|a_1|^2\left[\left|\pi_1(\cos\theta)\right|^2 + \left|\tau_1(\cos\theta)\right|^2\right] = \frac{3}{4}(1+\cos^2\theta)} \tag{7.48}$$

This means that, if the complex optical index contrast m is such that the term $\mathrm{Im}\left[(m^2-1)/(m^2+2)\right]$ differs sufficiently from zero, recalling that the condition

$$\mathrm{Im}\left(\frac{m^2-1}{m^2+2}\right) = 0$$

holds when m^2 is purely real, then the scattering efficiency is very small compared with the extinction efficiency, which itself becomes roughly equal to the absorption efficiency as a consequence.

7.3.4 Numerical Considerations

The Mie formulas (7.28), (7.39), (7.43), and (7.45), published in 1908, long remained unused due to the considerable difficulty involved in applying them to quantitative situations at the time. Today, this kind of problem has been largely overcome. However, the infinite series arising in these formulas are rather delicate, and programming must be done with great care.

In Chap. 4 of their book [3], Bohren and Huffman review the algorithms available in the 1980s for numerical calculation of the scattering coefficients a_n and b_n, and the extinction and scattering efficiencies $Q_{\mathrm{ext}}^{\mathrm{NP}}$ and $Q_{\mathrm{sca}}^{\mathrm{NP}}$, respectively. In general, the bigger the size parameter x, or the bigger its product with the modulus of the complex optical index contrast m, the more slowly the series converge. For example, an investigation of the scattering of visible light by a water droplet of diameter 1 mm (x and $|m|x$ of the order of 10^4) would require the calculation of several tens of thousands of the coefficients a_n and b_n in order to obtain a reasonable evaluation of the efficiencies $Q_{\mathrm{ext}}^{\mathrm{NP}}$ and $Q_{\mathrm{sca}}^{\mathrm{NP}}$. In 1996, W.J. Wiscombe [9] devised a calculation code called MIEV0 with a reputation for reliable results up to a size parameter of 20 000. However, the algorithm used by MIEV0 involves logarithmic derivatives

$$\mathrm{D}_n(\rho) = \frac{\mathrm{d}\left[\ln\left(\psi_n(\rho)\right)\right]}{\mathrm{d}\rho}$$

of the Ricatti–Bessel functions $\psi_n(\rho)$, and numerical calculation of these logarithmic derivatives turns out to be somewhat complicated. More recently, H. Du [10] has developed a new algorithm that does not require calculation of logarithmic derivatives, but involves instead the ratios

$$r_n(\rho) = \frac{\psi_{n-1}(\rho)}{\psi_n(\rho)}$$

of the functions $\psi_n(\rho)$. This particular routine is very easy to implement and has been tested successfully up to a value of 140 000 for the product $|m|x$.

7.3.5 Radiative Response of a Population of Spherical Particles

This point was already mentioned in Sect. 7.2.7. It was stressed that the transition from the radiative response of a single isolated particle to that of a whole population of particles is immediate, provided that the following three conditions are satisfied:

- The particles must be identical with regard to shape, size, and material content, and they must all be oriented in the same way.
- The particles must be randomly distributed in the host medium, with uniform probability density.
- The average distance between any two particles in the population must be big enough to ensure that the interaction between the electromagnetic field and any given particle occurs without any memory of previous interactions between the field and the other particles (the so-called independent radiation–matter interaction regime).

We shall henceforth focus on populations of spherical particles, assuming that they satisfy the second condition here, viz., random distribution within the host medium.

Let f be the volume fraction occupied by the particle population in the host medium. If f is small enough for the third condition to be satisfied, then it is a relatively simple matter to integrate a non-uniform distribution of particle radii in the particle population into the Mie calculations. To do this, rather than introducing a number of particles n per unit volume as in Sect. 7.2.7, one introduces a distribution $n(a)$ of radii present in the population, such that $n(a)\mathrm{d}a$ is the number of particles per unit volume with radii in the range a to $a+\mathrm{d}a$. Then the normalisation condition for the function $n(a)$ is

$$\int_0^\infty \frac{4\pi a^3}{3} n(a)\mathrm{d}a = f \ . \tag{7.49}$$

Having introduced this distribution $n(a)$, we can immediately write down expressions for the radiative properties β, σ, and $\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n})$ of the particle population.

Applying the principle of additivity of cross-sections which is valid in the independent matter–radiation interaction regime, this leads to the following expressions for the extinction and scattering coefficients β and σ:

$$\boxed{\beta = \int_0^\infty n(a) C_{\text{ext}}^{\text{NP}}(a)\mathrm{d}a\,, \qquad \sigma = \int_0^\infty n(a) C_{\text{sca}}^{\text{NP}}(a)\mathrm{d}a} \tag{7.50}$$

where $C_{\text{ext}}^{\text{NP}}(a)$ and $C_{\text{sca}}^{\text{NP}}(a)$ are the extinction and scattering cross-sections of a particle in the population with radius a. In the same way, summing the powers scattered into an elementary solid angle $\mathrm{d}\Omega$ encompassing a given direction $\mathbf{n}$ over all the intervals $[a, a+\mathrm{d}a]$ covering the radius distribution, the scattering phase function $\Phi^{\text{NP}}\big(\theta = (\mathbf{u},\mathbf{n})\big)$ of the particle population can be written in the form

$$\boxed{\Phi^{\text{NP}}(\theta) = \frac{1}{\sigma}\int_0^\infty n(a) C_{\text{sca}}^{\text{NP}}(a)\Phi^{\text{NP}}(a,\theta)\mathrm{d}a} \tag{7.51}$$

where $\Phi^{\text{NP}}(a,\theta)$ denotes the scattering phase function of a particle in the population with radius a.

As mentioned above, formulas (7.50) and (7.51), as well as (7.25) of Sect. 7.2.7 for a population of identical particles, were all obtained using the principle of additivity of the cross-sections. When the matter–radiation interaction regime can no longer be treated as independent, e.g., because the volume fraction of the particle population is such that the particles are very close to one another in comparison with the wavelength of the incident light, this additivity principle will no longer be justified. This raises the problem of determining what kind of matter–radiation interaction regime is applicable (independent or dependent) when radiation of given wavelength illuminates a population of particles with given size distribution and volume fraction.

In their monograph [11], C.L. Tien and B.L. Drolen begin to answer this question by means of a map of the (f,x) plane (f is the volume fraction and x the size parameter), dividing it into the dependent and independent matter–radiation interaction regimes (see Fig. 7.3). Their work concerns a population of spherical particles with uniform radius, dispersed randomly throughout the surrounding medium. The thick line on the map marks the boundary between the dependent scattering regime and the independent scattering regime, the latter being located in the region of small volume fractions and high size parameter. From the figure it can be seen that, as soon as the volume fraction f goes below 0.006, dependence effects can be ignored. The same is true as soon as the ratio c/λ goes above 0.5, where λ is the wavelength of the radiation and c is the average clearance between two particles within the population.

When the matter–radiation scattering regime is dependent, the radiative response of the particle population can no longer be deduced simply from the cross-sections of particles treated as isolated within the surrounding medium. In 1988, in their paper [12], H.S. Chu, A.J. Stretton, and C.L. Tien propose a theoretical expression for the ratio of the scattering efficiency $Q_{\text{sca}}^{\text{NP}}(\mathrm{D})$ in the dependent scattering regime to

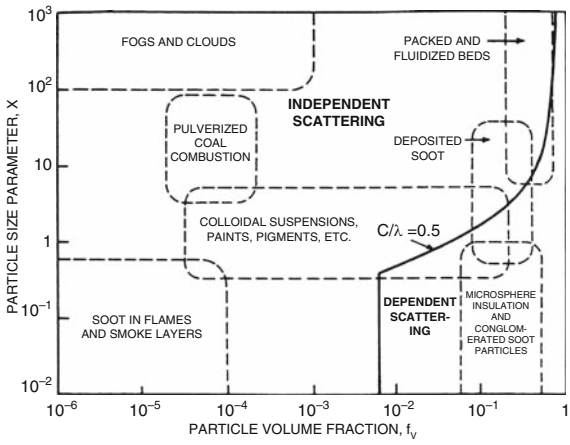

Fig. 7.3 Map of the (f,x) plane (f is the volume fraction and x the size parameter), showing the dependent and independent matter–radiation interaction regimes. Taken from [11]

its counterpart $Q_{\text{sca}}^{\text{NP}}(\text{I})$ in the independent scattering regime for a population of spherical particles with uniform radius, randomly dispersed throughout the surrounding medium:

$$\frac{Q_{\text{sca}}^{\text{NP}}(\text{D})}{Q_{\text{sca}}^{\text{NP}}(\text{I})} = \frac{(1-f)^4}{(1+2f)^2}\,. \tag{7.52}$$

Since the volume fraction f of the particle population is clearly less than or equal to 1, it follows that $Q_{\text{sca}}^{\text{NP}}(\text{D}) \leq Q_{\text{sca}}^{\text{NP}}(\text{I})$. Dependence effects are reflected in a reduction of the scattering efficiency which grows more significant as the volume fraction f increases. Furthermore, experimental work discussed in [12] shows that the trend observed here for the scattering efficiencies is exactly the opposite of what happens for the absorption efficiencies:

$$\frac{Q_{\text{abs}}^{\text{NP}}(\text{D})}{Q_{\text{abs}}^{\text{NP}}(\text{I})} \approx \left[\frac{Q_{\text{sca}}^{\text{NP}}(\text{D})}{Q_{\text{sca}}^{\text{NP}}(\text{I})}\right]^{-1} \approx \frac{(1+2f)^2}{(1-f)^4}\,. \tag{7.53}$$

In other words, a dependent matter–radiation scattering regime leads to two simultaneously occurring and antagonistic effects: a reduction in the overall efficiency of scattering on the one hand, and an increase in the overall efficiency of absorption on the other. The interested reader can find out more about this subject from the book by M. Kaviany [13] or the publication [14] by D. Baillis and J.-F. Sacadura. These two references review the state of the art regarding dependent scattering and associated effects.

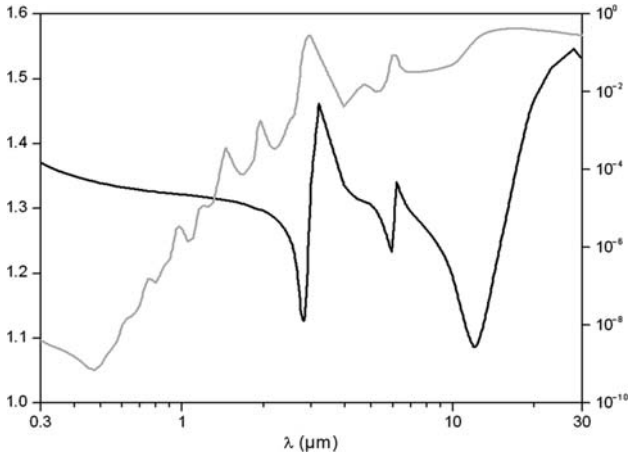

Fig. 7.4 Spectrum $m_\lambda = n_\lambda + \mathrm{i}\chi_\lambda$ of the complex optical index of water over the spectral range 0.3–30 μm. Data from [15]. *Black curve* and *left-hand scale*: Spectrum n_λ of the refractive index (no units). *Grey curve* and *right-hand scale*: Spectrum χ_λ of the extinction index (no units)

7.3.6 Application of Mie Theory to the Radiative Response of a Cloud

To end this section on Mie theory, we shall discuss the results of calculations based on this theory for a cloud made up of a population of water droplets suspended in the air (treated as a vacuum). The volume fraction of the droplets will be taken equal to 10^{-4}, which is therefore a rather dense cloud. Furthermore, the droplets making up the cloud will be assumed to have the same diameter of 1 μm. The Mie calculations are carried out for the spectral band 0.3–30 μm, covering the visible and near infrared range.

Since the radius of the droplets and their volume fraction f are given (0.5 μm and 10^{-4}, respectively), the only remaining piece of data required to carry out the calculation is the spectrum $m_\lambda = n_\lambda + \mathrm{i}\chi_\lambda$ of the complex optical index of water over the relevant wavelength band. We have taken this spectrum from the standard data base edited by E.D. Palik [15]. It is reproduced in Fig. 7.4. The main observations are that water is barely absorbent at all over the range 0.3–1.2 μm ($\chi_\lambda < 10^{-5}$ over this range), that there is a rather strong absorption peak at wavelength 2.95 μm ($\chi_\lambda = 0.28$ at this wavelength), and that it is rather absorbent between 4.4 and 30 μm ($\chi_\lambda > 10^{-2}$ over this range).

Figure 7.5 shows the spectra of the extinction coefficient β_λ and the scattering albedo ω_λ of the cloud as predicted by the Mie calculations. From 0.3 up to about 1.8 μm, the albedo ω_λ remains blocked at the value 1. Since the extinction index χ_λ of water is very small over this spectral range, the absorption phenomenon is completely dominated by scattering. This behaviour is suddenly reversed at the absorption peak mentioned above, located at wavelength 2.95 μm. This time, the spectrum ω_λ of the albedo exhibits a trough with minimal value 0.09, while the spectrum β_λ

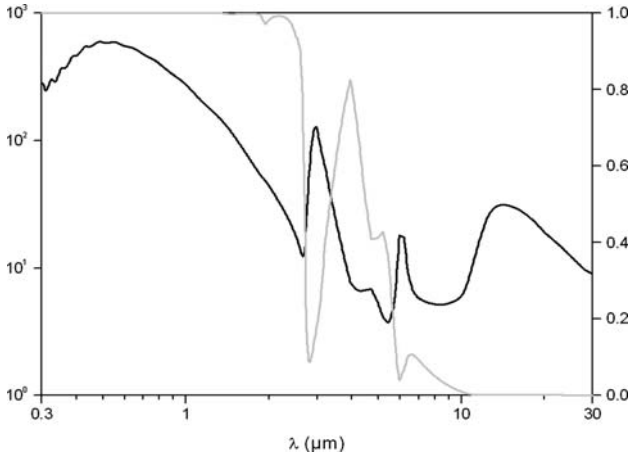

Fig. 7.5 Spectra of the radiative properties β_λ and ω_λ of the cloud resulting from Mie calculations. *Black curve* and *left-hand scale*: Spectrum β_λ of the extinction coefficient (expressed in m^{-1}). *Grey curve* and *right-hand scale*: Spectrum ω_λ of the scattering albedo (no units)

of the extinction coefficient has a peak with maximal value 130 m^{-1}. In the infrared part of the spectrum examined here, or more precisely, from about 4 μm, it is observed that the albedo ω_λ falls off steadily to reach 0 between 11 and 30 μm. There are two reasons for this decrease in the spectrum ω_λ :

- At long wavelengths, water becomes absorbent, because the values of its extinction index χ_λ become large.
- In this spectral range, the size parameter $x = 2\pi a/\lambda$ begins to get rather small (it is 0.29 at $\lambda = 11$ μm and 0.10 at $\lambda = 30$ μm), and in accordance with the results described in Sect. 7.3.3 when discussing Rayleigh scattering, the extinction efficiency (proportional to x) rapidly becomes dominant over the scattering efficiency (being proportional to x^4).

These two arguments together explain why absorption by the cloud controls extinction in the near infrared. To end here, we note that in the visible, the extinction coefficient β_λ remains within the range 250–600 m^{-1}. The characteristic extinction length, defined as the reciprocal of β_λ, is thus of the order of a few mm. So this really is a very dense cloud, and highly scattering in the visible.

The scattering phase functions predicted by the Mie theory in the interaction situation studied here also deserve some remarks. As the spectral range under investigation is 0.3–30 μm, the size parameter $x = 2\pi a/\lambda$ ranges over two orders of magnitude, from around 10 when $\lambda = 0.3$ μm to around 0.1 when $\lambda = 30$ μm, and it is interesting to examine to what extent the directional aspect of the scattering phenomenon is affected by the value of the parameter x. To answer this question, we calculated the scattering phase functions predicted by the Mie theory for three particular wavelengths, viz., $\lambda = 0.3$ μm (i.e., $x \approx 10$), $\lambda = 3$ μm (i.e., $x \approx 1$), and

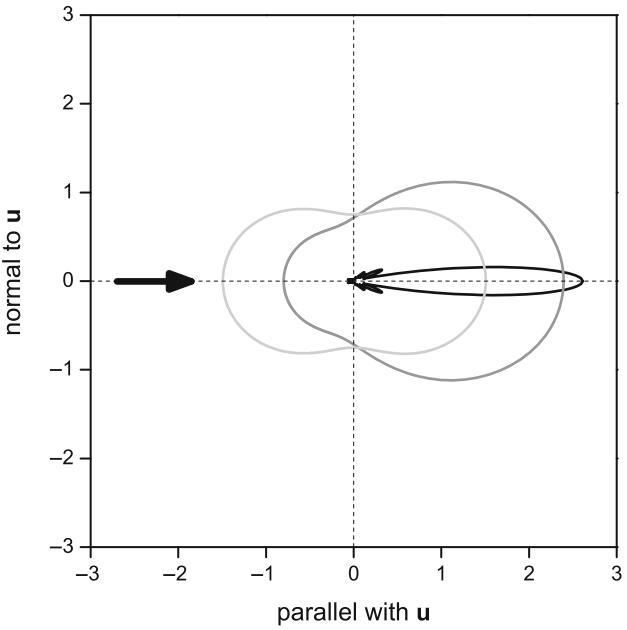

Fig. 7.6 Scattering phase functions plotted in the form of direction indicators, as predicted by Mie theory when electromagnetic radiation of wavelength λ interacts with a water droplet of diameter 1 μm. *Black indicator*: $\lambda = 0.3$ μm. *Dark grey indicator*: $\lambda = 3$ μm. *Light grey indicator*: $\lambda = 30$ μm. For clarity, the values of the phase function corresponding to the wavelength 0.3 μm have been divided by 20. The *black arrow* shows the direction **u** of the incident radiation

$\lambda = 30$ μm (i.e., $x \approx 0.1$). These three phase functions are shown in the form of direction indicators in Fig. 7.6. Note the following two points concerning the figure:

- For improved clarity, the values of the phase function corresponding to the wavelength 0.3 μm have been divided by 20.
- The black arrow indicates the direction **u** of the incident radiation.

A glance at the three phase functions plotted in Fig. 7.6 shows how far the directionality of the scattering phenomenon is affected by the value of the size parameter x. Whereas for $x \approx 0.1$, the phase function deviates only slightly from isotropy [indeed, it is very close to the Rayleigh scattering phase function expressed in (7.48)], it becomes extremely directional in the forward direction when x starts to grow. This well known property of Mie scattering phase functions is illustrated by other calculations in [6].

7.4 Discrete Dipole Approximation (DDA)

When the particles are not spherical, Mie theory can no longer treat the problem. This is also the case in situations where the particles can no longer be considered to be randomly dispersed throughout the host medium. So for example, when the

particles aggregate together to form a porous 3D material arrangement, Mie theory is no longer able to evaluate the bulk radiative response of these aggregates and must be replaced by another modelling technique able to integrate information concerning the relative positions of the particles within the aggregates.

The discrete dipole approximation (DDA), proposed in 1973 by E.M. Purcell and C.R. Pennypacker in what is unanimously agreed to be the founding article for this technique [16], is a very simple and flexible method for calculating the electromagnetic field scattered by a structured ensemble of polarisable volume elements with arbitrary spatial arrangement. The main developments occurred at the beginning of the 1990s in the hands of astrophysicists. For example, there are several applications of the DDA in the literature to calculate the matter–radiation interaction cross-sections of aggregates of interstellar dust [17–19]. B.T. Draine [19–21] made a major contribution to refining the theoretical aspects of the DDA and proposed a calculation code called DDSCAT, written in FORTRAN, restricted to an arrangement of dipoles at the nodes of a cubic lattice (freely available on the internet).

The discussion of the DDA to follow will be organised into three main parts. To begin with, we present the theory and limitations of the technique. We then describe various models of polarisability that can be envisaged as input data for the DDA calculations. Finally, we conclude by comparing the results of DDA and Mie calculations for two types of material structure.

7.4.1 The Theory of the DDA

The arguments justifying the mathematical formulas in [16, 19–21] are somewhat sketchy, and in addition are expressed using the CGS system. It thus seems useful to rewrite these demonstrations. The details can be found in [22], and we shall restrict here to a presentation of the main steps in the method of solution and a statement of the key formulas arising in the DDA.

Preamble: Radiation Emitted by an Oscillating Dipole

Consider a point dipole oscillating sinusoidally with angular frequency ω in the host medium (labelled 2). In complex notation, the dipole moment $\mathbf{P}(t)$ of the dipole can be written in the form $\mathbf{P}(t) = \boldsymbol{\Pi} \exp(-\mathrm{i}\omega t)$, where $\boldsymbol{\Pi}$ is a constant and possible complex vector. It is shown in [4] that the electromagnetic field emitted (or radiated) by the oscillating dipole has the general form

$$\boxed{\begin{cases} \mathbf{H}(\mathbf{r},t) = \dfrac{\mathrm{i}\omega}{4\pi}\dfrac{\exp(\mathrm{i}k_2 r)}{r^2}(1-\mathrm{i}k_2 r)\mathbf{n}\times\mathbf{P}(t)\,, \\ \mathbf{E}(\mathbf{r},t) = \dfrac{1}{4\pi\varepsilon_2}\dfrac{\exp(\mathrm{i}k_2 r)}{r^3}\Big\{(k_2 r)^2\Big[\big(\mathbf{n}\times\mathbf{P}(t)\big)\times\mathbf{n}\Big]+(1-\mathrm{i}k_2 r)\Big[3\big(\mathbf{n}\cdot\mathbf{P}(t)\big)\mathbf{n}-\mathbf{P}(t)\Big]\Big\}\,. \end{cases}} \tag{7.54}$$

Examination of these relations reveals that, in their general expression, the fields $\mathbf{H}(\mathbf{r},t)$ and $\mathbf{E}(\mathbf{r},t)$ vary spatially with terms in $1/r$ and $1/r^2$ for $\mathbf{H}(\mathbf{r},t)$, and $1/r$, $1/r^2$, and $1/r^3$ for $\mathbf{E}(\mathbf{r},t)$. However, if we are concerned with the radiative zone, i.e., the far-field region defined by $k_2 r \gg 1$, the expressions in (7.54) simplify to

$$\boxed{\begin{cases} \mathbf{H}(\mathbf{r},t) = \dfrac{\omega k_2}{4\pi}\dfrac{\exp(\mathrm{i}k_2 r)}{r}\mathbf{n}\times\mathbf{P}(t)\,, \\ \mathbf{E}(\mathbf{r},t) = \dfrac{k_2^2}{4\pi\varepsilon_2}\dfrac{\exp(\mathrm{i}k_2 r)}{r}\Big[\big(\mathbf{n}\times\mathbf{P}(t)\big)\times\mathbf{n}\Big] = \dfrac{k_2}{\omega\varepsilon_2}\Big[\mathbf{H}(\mathbf{r},t)\times\mathbf{n}\Big]\,. \end{cases}} \tag{7.55}$$

Finally, note that in the general expression (7.54) the electric field $\mathbf{E}(\mathbf{r},t)$ at a fixed point $\mathbf{r}$ is a linear and purely spatial operator, i.e., a function of $\mathbf{r}$ but not t, of the oscillating dipole moment $\mathbf{P}(t)$. This linearity relation can thus be written in the form

$$\boxed{\mathbf{E}(\mathbf{r},t) = \big[\mathsf{A}(\mathbf{r})\big]\bullet\mathbf{P}(t)} \tag{7.56}$$

where $\big[\mathsf{A}(\mathbf{r})\big]$ is a 3×3 matrix-valued function of $\mathbf{r}$, and the ensuing bullet denotes the matrix–vector product. Relative to a right-handed orthonormal basis $(\mathbf{x},\mathbf{y},\mathbf{z})$, it is easy to show that the matrix $\big[\mathsf{A}(\mathbf{r})\big]$ has the form

$$\boxed{\big[\mathsf{A}(\mathbf{r})\big] = \frac{\exp(\mathrm{i}k_2 r)}{4\pi\varepsilon_2 r^3}\left[(\rho^2+\mathrm{i}\rho-1)\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}-(\rho^2+3\mathrm{i}\rho-3)\begin{pmatrix}n_x^2&n_x n_y&n_x n_z\\n_x n_y&n_y^2&n_y n_z\\n_x n_z&n_y n_z&n_z^2\end{pmatrix}\right]} \tag{7.57}$$

where the dimensionless parameter ρ is defined by $\rho = k_2 r$, and n_x, n_y, n_z are the components of $\mathbf{n}$ in the basis $(\mathbf{x},\mathbf{y},\mathbf{z})$.

Assumptions and Limitations of the DDA

As mentioned above, the DDA is used to calculate the electromagnetic field scattered by a material object of arbitrary shape when it is illuminated by a plane incident monochromatic wave. Once the scattered electromagnetic field has been evaluated, one can then go on to calculate the radiative response of the material object in the form of the absorption, scattering, and extinction cross-sections, as well as the scattering phase function.

The basic idea of the DDA is to discretise the material object into volume elements that are small enough to be treated as oscillating dipoles induced partly by

the incident wave which activates them and partly by their pairwise interactions. The qualifier 'small enough' is to be understood here relative to the wavelength λ_0 of the incident radiation and also the local optical properties. A volume element resulting from the spatial discretisation will be considered small enough to be treated as an oscillating dipole if the following two conditions are satisfied:

- The path difference suffered by the incident wave when it crosses the element must be small compared with 2π. If this first condition is satisfied, then all points in the volume element can be considered to be excited with the same phase by the incident radiation. Let a be the characteristic dimension of the element. Then the condition just described takes the form

$$k_1 a = \frac{2\pi n_1 a}{\lambda_0} \ll 2\pi \,. \tag{7.58}$$

- When the material making up the element happens to be absorbent at the wavelength λ_0, the characteristic dimension of the element must also be small compared with the attenuation length of the incident wave in the material. This second requirement ensures that all points within the volume element will be excited with the same intensity by the incident radiation. It is expressed by the relation

$$\exp\left(-\frac{4\pi\chi_1 a}{\lambda_0}\right) \approx 1 \,, \quad \text{or} \quad \frac{4\pi\chi_1 a}{\lambda_0} \ll 1 \,, \tag{7.59}$$

  where χ_1 is the extinction index of the material making up the volume element, i.e., the imaginary part of its complex optical index m_1.

Draine [19] groups the two conditions (7.58) and (7.59) into one, written in the form

$$\boxed{|m_1| k_0 a < \beta} \tag{7.60}$$

where β is a constant of order unity. This condition imposes a maximum size on the volume elements produced by the spatial discretisation, and hence a minimal number of dipoles to ensure the validity of the discrete dipole approximation.

Note in passing that the notion of maximum size that can be attributed to a volume element is a local notion, because it does depend on local optical properties. As a consequence, if one hopes to apply the DDA to a heterogeneous material object with some position-dependent complex optical index m_1, it will be important to optimise the spatial discretisation, e.g., by ensuring that the volume elements are small at points where the condition (7.60) requires it, i.e., where $|m_1|$ is large, but not so small at points where it is not necessary, i.e., where $|m_1|$ is smaller.

Calculating the Induced Dipole Moments

As mentioned above, an oscillating dipole resulting from the spatial discretisation will be activated by the incident wave on the one hand, and by the electromagnetic

fields produced by neighbouring oscillating dipoles on the other. We begin therefore with the incident wave $\{\mathbf{E}_{\text{inc}}(\mathbf{r},t),\mathbf{H}_{\text{inc}}(\mathbf{r},t)\}$ given in (7.3). This wave activates the oscillating dipole moments at the centers of the N volume elements resulting from spatial discretisation. For any j between 1 and N, the oscillating dipole moment at the center $\mathbf{r}_j$ of the jth volume element can be expressed in the form $\mathbf{P}_j(t) = \boldsymbol{\Pi}_j \exp(-\mathrm{i}\omega t)$, where $\boldsymbol{\Pi}_j$ is a constant, complex vector to be determined. The vector $\boldsymbol{\Pi}_j$ is found by expressing the fact that the dipole moment $\mathbf{P}_j(t)$ is related to the local electric field $\mathbf{E}_{\text{loc}}(\mathbf{r}_j,t)$ at $\mathbf{r}_j$ through the relation $\mathbf{P}_j(t) = \alpha_j \mathbf{E}_{\text{loc}}(\mathbf{r}_j,t)$, where α_j is the polarisability of the jth volume element. Now the local electric field $\mathbf{E}_{\text{loc}}(\mathbf{r}_j,t)$ comprises the incident electric field $\mathbf{E}_{\text{inc}}(\mathbf{r}_j,t)$ and the electric fields produced by the neighbouring oscillating dipoles. This is expressed as follows:

$$\mathbf{E}_{\text{loc}}(\mathbf{r}_j,t) = \mathbf{E}_{\text{inc}}(\mathbf{r}_j,t) + \sum_{\substack{1\le k\le N\\ k\neq j}} \Big[\mathbf{E} \text{ produced at } \mathbf{r}_j \text{ by } \mathbf{P}_k(t) \text{ at } \mathbf{r}_k\Big] \,. \tag{7.61}$$

But a consequence of the result (7.56) is

$$\Big[\mathbf{E} \text{ produced at } \mathbf{r}_j \text{ by } \mathbf{P}_k(t) \text{ at } \mathbf{r}_k\Big] = \big[\mathsf{A}(\mathbf{r}_j - \mathbf{r}_k)\big] \bullet \mathbf{P}_k(t) \,, \tag{7.62}$$

whence (7.61) becomes

$$\mathbf{E}_{\text{loc}}(\mathbf{r}_j,t) = \mathbf{E}_{\text{inc}}(\mathbf{r}_j,t) + \sum_{\substack{1\le k\le N\\ k\neq j}} \big[\mathsf{A}(\mathbf{r}_j - \mathbf{r}_k)\big] \bullet \mathbf{P}_k(t) \,, \tag{7.63}$$

and the oscillating dipole moment $\mathbf{P}_j(t)$ must satisfy

$$\begin{aligned} \mathbf{P}_j(t) &= \alpha_j \left\{ \mathbf{E}_{\text{inc}}(\mathbf{r}_j,t) + \sum_{\substack{1\le k\le N\\ k\neq j}} \big[\mathsf{A}(\mathbf{r}_j - \mathbf{r}_k)\big] \bullet \mathbf{P}_k(t) \right\} \\ &= \alpha_j \left\{ E_0 \mathbf{e} \exp\Big[\mathrm{i}(k_2 \mathbf{u}\cdot\mathbf{r}_j - \omega t)\Big] + \sum_{\substack{1\le k\le N\\ k\neq j}} \big[\mathsf{A}(\mathbf{r}_j - \mathbf{r}_k)\big] \bullet \mathbf{P}_k(t) \right\} . \end{aligned} \tag{7.64}$$

Substituting the expressions $\mathbf{P}_j(t) = \boldsymbol{\Pi}_j \exp(-\mathrm{i}\omega t)$ and $\mathbf{P}_k(t) = \boldsymbol{\Pi}_k \exp(-\mathrm{i}\omega t)$ into this last equation, the terms in $\exp(-\mathrm{i}\omega t)$ simplify, and in the end the equations satisfied by the vectors $\boldsymbol{\Pi}_j$ take the form

$$\boxed{\frac{\boldsymbol{\Pi}_j}{\alpha_j} - \sum_{\substack{1\le k\le N\\ k\neq j}} \big[\mathsf{A}(\mathbf{r}_j - \mathbf{r}_k)\big] \bullet \boldsymbol{\Pi}_k = E_0 \mathbf{e} \exp(\mathrm{i}k_2 \mathbf{u}\cdot\mathbf{r}_j) \,, \quad \forall\, j \in \{1,2,\ldots,N\}} \tag{7.65}$$

This is a linear system of N vector equations in N unknown vectors. Moreover, the specific mathematical properties of the matrix operator $\big[\mathsf{A}(\mathbf{r})\big]$, a symmetric

operator satisfying $\big[\mathsf{A}(-\mathbf{r})\big] = \big[\mathsf{A}(\mathbf{r})\big]$, mean that the matrix for the linear system (7.65) is a symmetric complex-valued matrix. There thus exists a whole range of algorithms for solving the problem numerically and thereby evaluating the N vectors $\boldsymbol{\Pi}_j$.

Analysis of the linear system (7.65) shows that the vectors $\boldsymbol{\Pi}_j$ are proportional to the amplitude E_0 of the incident electric field, and also that, apart from this proportionality to E_0, they depend only on the dielectric permittivity ε_2 and the refractive index n_2 of the host medium, the position vectors $\mathbf{r}_k$ and polarisabilities α_k of all the oscillating dipoles, the direction of propagation $\mathbf{u}$ of the incident electromagnetic field, and its polarisation $\mathbf{e}$. Consequently, if the material configuration is fixed, so are the quantities ε_2, n_2, and all the $\mathbf{r}_k$ and α_k, and then the dipole moments $\boldsymbol{\Pi}_j$ are functions of the direction of propagation $\mathbf{u}$ and polarisation $\mathbf{e}$ alone, a result that can be expressed by writing $\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})$. Finally, note that the operator associating the whole set of dipole moments $\boldsymbol{\Pi}_j$ with the polarisation $\mathbf{e}$ is a linear operator.

Electromagnetic Field Scattered into the Far-Field Region by the Oscillating Dipole Ensemble

Once the dipole moments $\boldsymbol{\Pi}_j$ have been determined for a given pair $(\mathbf{u},\mathbf{e})$, one can calculate the electromagnetic field $\{\mathbf{E}_{\text{sca}}(\mathbf{r},t),\mathbf{H}_{\text{sca}}(\mathbf{r},t)\}$ radiated (or scattered) into the far field by the set of oscillating dipoles by summing over the individual contributions. The result (7.55) on p. 177 is used to write the total scattered electric and magnetic fields $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$ and $\mathbf{H}_{\text{sca}}(\mathbf{r},t)$ in the form

$$\begin{cases} \mathbf{H}_{\text{sca}}(\mathbf{r},t) = \dfrac{\omega k_2}{4\pi} \displaystyle\sum_{1\le j\le N} \frac{\exp\left(\mathrm{i}k_2\|\mathbf{r}-\mathbf{r}_j\|\right)}{\|\mathbf{r}-\mathbf{r}_j\|} \left[\frac{\mathbf{r}-\mathbf{r}_j}{\|\mathbf{r}-\mathbf{r}_j\|}\times\mathbf{P}_j(t)\right], \\ \mathbf{E}_{\text{sca}}(\mathbf{r},t) = \dfrac{k_2^2}{4\pi\varepsilon_2} \displaystyle\sum_{1\le j\le N} \frac{\exp\left(\mathrm{i}k_2\|\mathbf{r}-\mathbf{r}_j\|\right)}{\|\mathbf{r}-\mathbf{r}_j\|} \left\{\left[\frac{\mathbf{r}-\mathbf{r}_j}{\|\mathbf{r}-\mathbf{r}_j\|}\times\mathbf{P}_j(t)\right]\times\frac{\mathbf{r}-\mathbf{r}_j}{\|\mathbf{r}-\mathbf{r}_j\|}\right\}. \end{cases} \tag{7.66}$$

Now it is easy to show that, when $\|\mathbf{r}\|\to\infty$ (the far-field hypothesis), an expansion to first order of $\|\mathbf{r}-\mathbf{r}_j\|$ gives $r-\mathbf{n}\cdot\mathbf{r}_j$ (recalling the notation $r=\|\mathbf{r}\|$ and $\mathbf{n}=\mathbf{r}/r$). Using this approximation, the expression for the scattered magnetic field in (7.66) becomes, in the far field,

$$\boxed{\mathbf{H}_{\text{sca}}(\mathbf{r},t) \approx \frac{\omega k_2}{4\pi}\frac{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]}{r}\left\{\mathbf{n}\times\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j\exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\}} \tag{7.67}$$

By the same process, it is easy to show that the electric field $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$ scattered into the far field by the ensemble of oscillating dipoles has the form

$$\boxed{\mathbf{E}_{\rm sca}(\mathbf{r},t) \approx \frac{k_2^2}{4\pi\varepsilon_2}\frac{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]}{r}\left\{\mathbf{n}\times\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j \exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\}\times\mathbf{n}} \tag{7.68}$$

In the final expressions (7.67) and (7.68) for the fields $\mathbf{H}_{\rm sca}(\mathbf{r},t)$ and $\mathbf{E}_{\rm sca}(\mathbf{r},t)$, there appears a common vector

$$\boxed{\boldsymbol{\Theta} = \sum_{1\le j\le N}\boldsymbol{\Pi}_j \exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)}$$

which is a function of the N dipole moments $\boldsymbol{\Pi}_j$. Since the material configuration is considered fixed, it is a function only of the three vectors $\mathbf{u}$, $\mathbf{e}$, and $\mathbf{n}$, i.e., it has the form $\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{n})$. The fields $\mathbf{H}_{\rm sca}(\mathbf{r},t)$ and $\mathbf{E}_{\rm sca}(\mathbf{r},t)$ are expressed in terms of the vector $\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{n})$ via the relations

$$\boxed{\begin{cases}\mathbf{H}_{\rm sca}(\mathbf{r},t) = \dfrac{\omega k_2}{4\pi}\dfrac{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]}{r}\left[\mathbf{n}\times\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{n})\right]\,, \\ \mathbf{E}_{\rm sca}(\mathbf{r},t) = \dfrac{k_2^2}{4\pi\varepsilon_2}\dfrac{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]}{r}\left\{\left[\mathbf{n}\times\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{n})\right]\times\mathbf{n}\right\} = \dfrac{k_2}{\omega\varepsilon_2}\mathbf{H}_{\rm sca}(\mathbf{r},t)\times\mathbf{n}\,.\end{cases}} \tag{7.69}$$

These expressions look very similar to those in (7.55) for the electromagnetic field produced by a single oscillating dipole, except for the significant difference that, in the present context, the vector playing the role for the dipole moment, viz., $\boldsymbol{\Theta}$, depends on the scattering direction $\mathbf{n}$.

Having found the electromagnetic field scattered into the far field by the ensemble of N oscillating dipoles, we may now pursue our analysis by expressing the radiative response, i.e., the absorption, scattering, and extinction cross-sections and the scattering phase function, of the ensemble of N dipoles using the general results presented in Sect. 7.2. This is the task to which we shall now turn.

Extinction and Scattering Cross-Sections for the Oscillating Dipole Ensemble

In Sect. 7.2.4 we wrote the extinction and scattered powers $\Psi^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e})$ and $\Psi^{\rm P}_{\rm sca}(\mathbf{u},\mathbf{e})$ under polarised illumination in the form (7.14) and (7.19), respectively, viz.,

$$\Psi^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e}) = \frac{2\pi\varepsilon_2 c_2}{k_2^2}\mathrm{Re}\left\{\left[\left(\frac{\mathbf{E}_{\rm inc}(\mathbf{r},t)}{\exp\left[\mathrm{i}(k_2\mathbf{u}\cdot\mathbf{r}-\omega t)\right]}\right)^*\cdot\left(\frac{\mathbf{E}_{\rm sca}(\mathbf{r},t)}{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]/(-\mathrm{i}k_2 r)}\right)\right]_{\substack{r\to+\infty\\ \theta=0}}\right\}, \tag{7.70}$$

$$\Psi^{\rm P}_{\rm sca}(\mathbf{u},\mathbf{e}) = \frac{1}{2\omega\mu_2 k_2}\int_{4\pi}\left\|\mathbf{A}_{\rm sca}(\mathbf{n})\right\|^2 \mathrm{d}\Omega\,. \tag{7.71}$$

The amplitude vector field $\mathbf{A}_{\rm sca}(\mathbf{n})$ was defined in (7.15). Since $\mathbf{E}_{\rm sca}(\mathbf{r},t)$ is now known in the far field [see (7.69)], the expression for $\mathbf{A}_{\rm sca}(\mathbf{n})$ follows immediately. Further, the term

$$\left[\frac{\mathbf{E}_{\rm sca}(\mathbf{r},t)}{\exp\left[\mathrm{i}(k_2 r-\omega t)\right]/(-\mathrm{i}k_2 r)}\right]_{\substack{r\to+\infty\\ \theta=0}}$$

appearing in the expression for $\Psi^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e})$ must be evaluated in the far field ($r\to\infty$). It is thus exactly equal to $\mathbf{A}_{\rm sca}(\mathbf{n})$. Finally, this same term must be evaluated in the forward direction ($\theta=0$) for which the scattering direction $\mathbf{n}$ is equal to the direction of incidence $\mathbf{u}$. It is thus equal to $\mathbf{A}_{\rm sca}(\mathbf{u})$. Taking these points into account, (7.70) for $\Psi^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e})$ becomes

$$\Psi^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e})=\frac{2\pi\varepsilon_2 c_2}{k_2^2}\mathrm{Re}\left[(E_0\mathbf{e})^*\cdot\left\{-\frac{\mathrm{i}k_2^3}{4\pi\varepsilon_2}\left[\mathbf{u}\times\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{u})\right]\times\mathbf{u}\right\}\right]. \tag{7.72}$$

After some manipulation and using the fact that $\mathbf{e}\perp\mathbf{u}$, this can be written in the form

$$\begin{aligned}\Psi^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e}) &= \frac{\omega E_0}{2}\mathrm{Im}\left[\mathbf{e}^*\cdot\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{u})\right]\\ &= \frac{\omega E_0}{2}\mathrm{Im}\left\{\mathbf{e}^*\cdot\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})\exp(-\mathrm{i}k_2\mathbf{u}\cdot\mathbf{r}_j)\right]\right\}.\end{aligned} \tag{7.73}$$

Dividing both sides of the last equation by the incident power per unit area $\psi_{\rm inc}$ given by (7.7), we finally arrive at an expression for the extinction cross-section under polarised radiation:

$$\boxed{C^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e})=\frac{\Psi^{\rm P}_{\rm ext}(\mathbf{u},\mathbf{e})}{\psi_{\rm inc}}=\frac{k_2}{\varepsilon_2 E_0}\mathrm{Im}\left\{\mathbf{e}^*\cdot\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})\exp(-\mathrm{i}k_2\mathbf{u}\cdot\mathbf{r}_j)\right]\right\}} \tag{7.74}$$

An expression for the extinction cross-section under non-polarised illumination $C^{\rm NP}_{\rm ext}(\mathbf{u})$ can then be deduced immediately from the conclusions of Sect. 7.2.6:

$$\boxed{C^{\rm NP}_{\rm ext}(\mathbf{u})=\frac{k_2}{2\varepsilon_2 E_0}\mathrm{Im}\left\{\sum_{1\le j\le N}\left[\mathbf{v}\cdot\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{v})+\mathbf{w}\cdot\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{w})\right]\exp(-\mathrm{i}k_2\mathbf{u}\cdot\mathbf{r}_j)\right\}} \tag{7.75}$$

where the vectors $\mathbf{v}$ and $\mathbf{w}$ are two real vectors making up an orthonormal basis with $\mathbf{u}$.

Let us now consider the scattered power $\Psi^{\rm P}_{\rm sca}(\mathbf{u},\mathbf{e})$ under polarised illumination. As the expression for the amplitude vector field $\mathbf{A}_{\rm sca}(\mathbf{n})$ is now known, (7.71) for $\Psi^{\rm P}_{\rm sca}(\mathbf{u},\mathbf{e})$ transforms to

$$\Psi_{\rm sca}^{\rm P}(\mathbf{u},\mathbf{e}) = \frac{c_2 k_2^4}{32\pi^2\varepsilon_2}\int_{4\pi}\Big\|\big[\mathbf{n}\times\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{n})\big]\times\mathbf{n}\Big\|^2 {\rm d}\Omega\,, \tag{7.76}$$

and after straightforward mathematical manipulation, this becomes

$$\begin{aligned}\Psi_{\rm sca}^{\rm P}(\mathbf{u},\mathbf{e}) &= \frac{c_2 k_2^4}{32\pi^2\varepsilon_2}\int_{4\pi}\Big\|\mathbf{n}\times\boldsymbol{\Theta}(\mathbf{u},\mathbf{e},\mathbf{n})\Big\|^2 {\rm d}\Omega \\ &= \frac{c_2 k_2^4}{32\pi^2\varepsilon_2}\int_{4\pi}\left\|\mathbf{n}\times\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})\exp(-{\rm i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2 {\rm d}\Omega\,. \end{aligned}\tag{7.77}$$

After dividing by the incident power per unit area $\psi_{\rm inc}$, we obtain an expression for the scattering cross-section under polarised radiation:

$$\boxed{C_{\rm sca}^{\rm P}(\mathbf{u},\mathbf{e}) = \frac{\Psi_{\rm sca}^{\rm P}(\mathbf{u},\mathbf{e})}{\psi_{\rm inc}} = \left(\frac{k_2^2}{4\pi\varepsilon_2 E_0}\right)^2\int_{4\pi}\left\|\mathbf{n}\times\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})\exp(-{\rm i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2 {\rm d}\Omega} \tag{7.78}$$

Finally, we derive the scattering cross-section under non-polarised radiation:

$$\boxed{\begin{aligned}C_{\rm sca}^{\rm NP}(\mathbf{u}) = \frac{1}{2}\left(\frac{k_2^2}{4\pi\varepsilon_2 E_0}\right)^2\int_{4\pi}\Bigg\{&\left\|\mathbf{n}\times\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{v})\exp(-{\rm i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2 \\ &+\left\|\mathbf{n}\times\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{w})\exp(-{\rm i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2\Bigg\}{\rm d}\Omega\end{aligned}} \tag{7.79}$$

where the vectors $\mathbf{v}$ and $\mathbf{w}$ were introduced above. The directional integrals appearing in expressions (7.76)–(7.79) for the scattered power $\Psi_{\rm sca}^{\rm P}(\mathbf{u},\mathbf{e})$ and the scattering cross-sections $C_{\rm sca}^{\rm P}(\mathbf{u},\mathbf{e})$ and $C_{\rm sca}^{\rm NP}(\mathbf{u})$ can be calculated numerically, using the discrete ordinate methods, for example. These techniques are described in [23].

Scattering Phase Functions for the Oscillating Dipole Ensemble

The expressions for these phase functions under polarised and non-polarised illumination, viz., $\Phi^{\rm P}(\mathbf{u},\mathbf{e},\mathbf{n})$ and $\Phi^{\rm NP}(\mathbf{u},\mathbf{n})$, follow directly from (7.78) and (7.79) above. Up to multiplicative constants, they are given respectively by

$$\left\|\mathbf{n}\times\left[\sum_{1\le j\le N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})\exp(-{\rm i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2 \tag{7.80}$$

and

$$\left\|\mathbf{n}\times\left[\sum_{1\leq j\leq N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{v})\exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2+\left\|\mathbf{n}\times\left[\sum_{1\leq j\leq N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{w})\exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2, \tag{7.81}$$

where the multiplicative constants are such that the functions $\boldsymbol{\Phi}^{\mathrm{P}}(\mathbf{u},\mathbf{e},\mathbf{n})$ and $\boldsymbol{\Phi}^{\mathrm{NP}}(\mathbf{u},\mathbf{n})$ satisfy the normalisation condition

$$\int_{4\pi}\boldsymbol{\Phi}(\mathbf{n})\mathrm{d}\Omega=4\pi\,.$$

This leads to the final expressions

$$\boxed{\boldsymbol{\Phi}^{\mathrm{P}}(\mathbf{u},\mathbf{e},\mathbf{n})=\frac{4\pi}{C^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{e})}\left(\frac{k_2^2}{4\pi\varepsilon_2E_0}\right)^2\left\|\mathbf{n}\times\left[\sum_{1\leq j\leq N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})\exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2} \tag{7.82}$$

and

$$\boxed{\begin{aligned}\boldsymbol{\Phi}^{\mathrm{NP}}(\mathbf{u},\mathbf{n})=\frac{2\pi}{C^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u})}\left(\frac{k_2^2}{4\pi\varepsilon_2E_0}\right)^2\Bigg\{&\left\|\mathbf{n}\times\left[\sum_{1\leq j\leq N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{v})\exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2\\&+\left\|\mathbf{n}\times\left[\sum_{1\leq j\leq N}\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{w})\exp(-\mathrm{i}k_2\mathbf{n}\cdot\mathbf{r}_j)\right]\right\|^2\Bigg\}\end{aligned}} \tag{7.83}$$

Absorption Cross-Sections for the Oscillating Dipole Ensemble

Up to now, the problem of calculating the absorption cross-section $C^{\mathrm{P}}_{\mathrm{abs}}(\mathbf{u},\mathbf{e})$ for a given direction of propagation $\mathbf{u}$ and polarisation $\mathbf{e}$ has been left to one side. More precisely, no alternative to the simple subtraction

$$C^{\mathrm{P}}_{\mathrm{abs}}(\mathbf{u},\mathbf{e})=C^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{e})-C^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{e})$$

has yet been put forward. However, in the specific context of Mie theory, the two sequences of coefficients a_n and b_n characterising the scattered electromagnetic field $\left\{\mathbf{E}_{\mathrm{sca}}(r,\theta,\varphi),\mathbf{H}_{\mathrm{sca}}(r,\theta,\varphi)\right\}$ have been expressed analytically [see (7.28) in Sect. 7.3.1], so in principle there is no reason why the absorbed power $\Psi^{\mathrm{P}}_{\mathrm{abs}}(\mathbf{u},\mathbf{e})$ should not be evaluated using the general method described in Sect. 7.2.3. The reason why no Mie formula exists for the absorption phenomenon is very likely because the calculations are somewhat involved for this particular phenomenon. We shall see in the following that, in contrast to Mie theory, using the dipole moments $\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})$, previously evaluated by means of the DDA by solving the linear system (7.65), the cross-section $C^{\mathrm{P}}_{\mathrm{abs}}(\mathbf{u},\mathbf{e})$ can be independently calculated. It is then possible to estimate the quality and consistency of the numerical calculations by comparing the sum $C^{\mathrm{P}}_{\mathrm{abs}}(\mathbf{u},\mathbf{e})+C^{\mathrm{P}}_{\mathrm{sca}}(\mathbf{u},\mathbf{e})$ with the cross-section $C^{\mathrm{P}}_{\mathrm{ext}}(\mathbf{u},\mathbf{e})$.

The method for evaluating the absorption cross-section C_{abs} using the dipole moments $\boldsymbol{\Pi}_j$ is mentioned in [16] and described in more detail in [19]. The idea is to begin by isolating in space one of the N oscillating dipoles from the DDA calculation and to express the absorption cross-section $(c_{\text{abs}})_j$ of this particular jth dipole using the relation $(c_{\text{abs}})_j = (c_{\text{ext}})_j - (c_{\text{sca}})_j$ (with the obvious notation), and taking care to remove all trace of the incident electric field from this expression to the benefit of the dipole moment $\mathbf{P}_j(t) = \boldsymbol{\Pi}_j \exp(-\mathrm{i}\omega t)$. As this dipole moment $\mathbf{P}_j(t)$ is generated by a DDA calculation, it accounts for the interactions between the jth dipole and its $N-1$ counterparts, so the sum of the elementary absorption cross-sections in $\sum_{1\leq j\leq N}(c_{\text{abs}})_j$ produces a total absorption cross-section C_{abs} that is representative of the absorption by the whole ensemble of N oscillating dipoles. Now that we have described this idea for calculating the absorption cross-section C_{abs}, it is important to emphasise that the relation $C = \sum_{1\leq j\leq N} c_j$ mentioned above is only valid for the absorption. Indeed, while absorption is additive in energy terms, scattering is not, and neither therefore is extinction.

We thus consider an isolated oscillating point dipole, located at $\mathbf{r} = \mathbf{0}$, with dipole moment $\mathbf{P}(t) = \boldsymbol{\Pi} \exp(-\mathrm{i}\omega t)$ activated by polarised incident radiation in which the electric field is given by

$$\mathbf{E}_{\text{inc}}(\mathbf{r},t) = E_0\mathbf{e} \exp\left[\mathrm{i}(k_2\mathbf{u}\cdot\mathbf{r} - \omega t)\right] .$$

Since this dipole is assumed to be alone in space, its dipole moment is $\mathbf{P}(t) = \alpha\mathbf{E}_{\text{inc}}(\mathbf{r} = \mathbf{0}, t)$, whence $E_0\mathbf{e} = \boldsymbol{\Pi}/\alpha$.

As we saw on p. 177, this dipole scatters a field $\left\{\mathbf{E}_{\text{sca}}(\mathbf{r},t), \mathbf{H}_{\text{sca}}(\mathbf{r},t)\right\}$ given in the far field by (7.55), viz.,

$$\begin{cases} \mathbf{H}_{\text{sca}}(\mathbf{r},t) = \dfrac{\omega k_2}{4\pi}\dfrac{\exp(\mathrm{i}k_2 r)}{r}\mathbf{n}\times\mathbf{P}(t) , \\ \mathbf{E}_{\text{sca}}(\mathbf{r},t) = \dfrac{k_2^2}{4\pi\varepsilon_2}\dfrac{\exp(\mathrm{i}k_2 r)}{r}\Big[\big(\mathbf{n}\times\mathbf{P}(t)\big)\times\mathbf{n}\Big] = \dfrac{k_2}{\omega\varepsilon_2}\Big[\mathbf{H}_{\text{sca}}(\mathbf{r},t)\times\mathbf{n}\Big] . \end{cases} \tag{7.84}$$

The extinction cross-section $c_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{e})$ of this dipole is calculated using the optical theorem (see Sect. 7.2.4). Starting with the general expression (7.14) for the extinction power and the relation (7.15) defining the amplitude vector field $\mathbf{A}_{\text{sca}}(\mathbf{n})$, not forgetting that in the forward direction, i.e., for $\theta = 0$, the scattering direction $\mathbf{n}$ is equal to the illumination direction $\mathbf{u}$, and replacing the term $E_0\mathbf{e}$ in the expression for the incident electric field $\mathbf{E}_{\text{inc}}(\mathbf{r},t)$ by $\boldsymbol{\Pi}/\alpha$, straightforward but tedious calculation leads to the following expression for the cross-section $c_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{e})$:

$$c_{\text{ext}}^{\text{P}}(\mathbf{u},\mathbf{e}) = \frac{k_2}{\varepsilon_2 E_0^2}\mathrm{Im}\left\{\frac{\boldsymbol{\Pi}^*}{\alpha^*}\cdot\big[(\mathbf{u}\times\boldsymbol{\Pi})\times\mathbf{u}\big]\right\} . \tag{7.85}$$

Moreover, since $\boldsymbol{\Pi}$ is in this case equal to $\alpha E_0\mathbf{e}$, it follows that $\boldsymbol{\Pi} \perp \mathbf{u}$ and hence $(\mathbf{u}\times\boldsymbol{\Pi})\times\mathbf{u} = \boldsymbol{\Pi}$. Equation (7.85) then simplifies to

$$\boxed{c_{\mathrm{ext}}^{\mathrm{P}}(\mathbf{u},\mathbf{e}) = \frac{k_2 \left\| \boldsymbol{\Pi} \right\|^2}{\varepsilon_2 E_0^2} \mathrm{Im} \frac{1}{\alpha^*}} \tag{7.86}$$

The elementary scattering cross-section $c_{\mathrm{sca}}^{\mathrm{P}}(\mathbf{u},\mathbf{e})$ is determined along the same lines. The scattered power is calculated using (7.19). Since the expression for the amplitude vector field $\mathbf{A}_{\mathrm{sca}}(\mathbf{n})$ is known, we begin by deducing the following expression for the cross-section $c_{\mathrm{sca}}^{\mathrm{P}}(\mathbf{u},\mathbf{e})$:

$$c_{\mathrm{sca}}^{\mathrm{P}}(\mathbf{u},\mathbf{e}) = \left(\frac{k_2^2}{4\pi\varepsilon_2 E_0} \right)^2 \int_{4\pi} \left\| (\mathbf{n}\times\boldsymbol{\Pi})\times\mathbf{n} \right\|^2 \mathrm{d}\Omega \, . \tag{7.87}$$

Then noting that $\left\| (\mathbf{n}\times\boldsymbol{\Pi})\times\mathbf{n} \right\|^2 = \left\| \mathbf{n}\times\boldsymbol{\Pi} \right\|^2$, since the vector $\mathbf{n}$ is real and unit length, and noting also that

$$\int_{4\pi} \left\| \mathbf{n}\times\boldsymbol{\Pi} \right\|^2 \mathrm{d}\Omega = \frac{8\pi \left\| \boldsymbol{\Pi} \right\|^2}{3} \, ,$$

since $\boldsymbol{\Pi}$ is a constant vector, the expression (7.87) simplifies to

$$\boxed{c_{\mathrm{sca}}^{\mathrm{P}}(\mathbf{u},\mathbf{e}) = \frac{k_2^4 \left\| \boldsymbol{\Pi} \right\|^2}{6\pi\varepsilon_2^2 E_0^2}} \tag{7.88}$$

The absorption cross-section $c_{\mathrm{abs}}^{\mathrm{P}}(\mathbf{u},\mathbf{e})$ of an isolated dipole is thus expressed solely as a function of the polarisability α and the dipole moment $\boldsymbol{\Pi}$ according to the relation

$$\boxed{c_{\mathrm{abs}}^{\mathrm{P}}(\mathbf{u},\mathbf{e}) = c_{\mathrm{ext}}^{\mathrm{P}}(\mathbf{u},\mathbf{e}) - c_{\mathrm{sca}}^{\mathrm{P}}(\mathbf{u},\mathbf{e}) = \frac{k_2 \left\| \boldsymbol{\Pi} \right\|^2}{\varepsilon_2 E_0^2} \left(\mathrm{Im} \frac{1}{\alpha^*} - \frac{k_2^3}{6\pi\varepsilon_2} \right)} \tag{7.89}$$

If we now return to our DDA calculations applied to an ensemble of N oscillating dipoles, the absorption cross-section $C_{\mathrm{abs}}^{\mathrm{P}}(\mathbf{u},\mathbf{e})$ of this ensemble of dipoles can be evaluated with the help of the additivity principle mentioned earlier:

$$\boxed{C_{\mathrm{abs}}^{\mathrm{P}}(\mathbf{u},\mathbf{e}) = \sum_{1\le j\le N} \left(c_{\mathrm{abs}}^{\mathrm{P}} \right)_j (\mathbf{u},\mathbf{e}) = \frac{k_2}{\varepsilon_2 E_0^2} \sum_{1\le j\le N} \left\| \boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e}) \right\|^2 \left(\mathrm{Im} \frac{1}{\alpha_j^*} - \frac{k_2^3}{6\pi\varepsilon_2} \right)} \tag{7.90}$$

Underlying the argument just presented was the assumption of polarised incident illumination, with direction of propagation $\mathbf{u}$ and polarisation $\mathbf{e}$. The expression (7.90) for the absorption cross-section $C_{\mathrm{abs}}^{\mathrm{P}}(\mathbf{u},\mathbf{e})$ thus involves the dipole moments $\boldsymbol{\Pi}_j(\mathbf{u},\mathbf{e})$ in this particular illumination configuration. When the incident radiation is not polarised, the absorption cross-section $C_{\mathrm{abs}}^{\mathrm{NP}}(\mathbf{u})$ is given by

$$\boxed{C_{\text{abs}}^{\text{NP}}(\mathbf{u}) = \frac{k_2}{2\varepsilon_2 E_0^2} \sum_{1\le j\le N} \left[\left\| \boldsymbol{\Pi}_j(\mathbf{u},\mathbf{v}) \right\|^2 + \left\| \boldsymbol{\Pi}_j(\mathbf{u},\mathbf{w}) \right\|^2 \right] \left(\text{Im}\frac{1}{\alpha_j^*} - \frac{k_2^3}{6\pi\varepsilon_2} \right)} \tag{7.91}$$

which only depends on the direction of incidence $\mathbf{u}$.

7.4.2 Models for Polarisability

The linear system (7.65) underlying the discrete dipole approximation naturally brings in the polarisabilities α_j of the N oscillating dipoles. We must therefore determine these N values α_j before beginning any DDA calculation on the population.

The Clausius–Mossotti Formulation

The model generally taken as the starting point for calculating the polarisability of a material element is the Clausius–Mossotti model [3, 16, 19], which follows from a standard result of electrostatics concerning the interaction between a uniform electric field and a spherical particle (labelled 1) immersed in a host medium (labelled 2), assumed infinite. Let $E_0\mathbf{x}$ be this field when there is no particle. Introducing standard polar coordinates (r,θ) with an arbitrary origin O in the host medium [$r = \|\mathbf{r}\|$ and $\theta = (\mathbf{x},\mathbf{r})$], the potential field $V(\mathbf{r})$ associated with the electric field $E_0\mathbf{x}$ is given by $V(\mathbf{r}) = -E_0 x = -E_0 r\cos\theta$. If the particle is now brought to O, the potential field $V(\mathbf{r})$ will be affected by the presence of the particle. In particular, it can be shown [3] that this field $V(\mathbf{r})$ has the following analytic expression outside the particle, i.e., in the host medium:

$$V(\mathbf{r}) = -E_0 r\cos\theta + a^3 E_0 \frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+2\varepsilon_2}\frac{\cos\theta}{r^2} = -E_0 x + a^3 E_0 \frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+2\varepsilon_2}\frac{\cos\theta}{r^2}\,, \tag{7.92}$$

where ε_1 and ε_2 are the static dielectric permittivities of the particle and the host medium, respectively, and a is the radius of the particle. Calculating minus the gradient of the expression in (7.92), we immediately reach the following conclusion: the electric field $\mathbf{E}(\mathbf{r})$ in the host medium in the presence of the particle is equal to the sum of the uniform field $E_0\mathbf{x}$ that would exist there in the absence of the particle and a supplementary field resulting from the perturbation due to the particle in the host medium, where the latter is given by

$$-a^3 E_0 \frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+2\varepsilon_2}\nabla\left(\frac{\cos\theta}{r^2}\right).$$

Clearly, the uniform field $E_0\mathbf{x}$ and the supplementary field just displayed are the static limits of the incident electric field $\mathbf{E}_{\text{inc}}(\mathbf{r},t)$ and the scattered electric field $\mathbf{E}_{\text{sca}}(\mathbf{r},t)$ of the electromagnetic Mie theory. Furthermore, returning to the potential

field $V(\mathbf{r})$ given in (7.92), we observe that its supplementary term

$$a^3 E_0 \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2} \frac{\cos\theta}{r^2}$$

which reflects the presence of the particle and the perturbation it causes within the host medium, looks very similar to the expression for the potential produced by an electrostatic dipole. Indeed, one can prove the following standard result:

$$V(\mathbf{r}) = \frac{P}{4\pi\varepsilon_2} \frac{\cos\theta}{r^2} \,, \tag{7.93}$$

for an electrostatic dipole of moment $P\mathbf{x}$ placed in a host medium with static dielectric permittivity ε_2. Hence the supplementary potential

$$a^3 E_0 \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2} \frac{\cos\theta}{r^2}$$

appearing in (7.92) can be interpreted as the potential produced in the host medium by an electrostatic dipole of moment

$$\mathbf{P} = 4\pi\varepsilon_2 a^3 E_0 \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2} \mathbf{x} \,. \tag{7.94}$$

In other words, from the point of view of the potential field produced in the host medium, the spherical particle of radius a and permittivity ε_1 behaves exactly like an electrostatic dipole of moment $\mathbf{P}$ as defined in (7.94). As this dipole moment $\mathbf{P}$ is proportional to the exciting electric field $E_0\mathbf{x}$, the polarisability α is defined as the coefficient of proportionality between $\mathbf{P}$ and $E_0\mathbf{x}$, viz.,

$$\boxed{\mathbf{P} = \alpha E_0 \mathbf{x} \,, \quad \alpha = 4\pi\varepsilon_2 a^3 \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2}} \tag{7.95}$$

Introducing the volume $v = 4\pi a^3/3$ of the spherical particle, the expression for α in (7.95) can be rewritten in the form

$$\boxed{\alpha = 3\varepsilon_2 v \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2}} \tag{7.96}$$

The above relation is commonly called the Clausius–Mossotti relation. Strictly speaking, it is only applicable to the static polarisability of a spherical particle. In practice, however, it is commonly accepted that (7.96) can be extended to the case of a particle with arbitrary shape (and volume v) and to calculations of monochromatic polarisability (the quantities ε_1 and ε_2 then refer to the monochromatic dielectric permittivities of the particle and host medium, respectively).

Other Formulations of Polarisability

Several authors have put forward refined expressions for the polarisability α. A review of these different formulations is presented in the first part of [20]. In the following, we shall quote two particular expressions we have implemented in investigations carried out at CEA/Le Ripault (France): the formulations by Draine [19] and Doyle [24].

Draine Formulation

In his paper [19], the author points out that, although the Clausius–Mossotti formula is exact in the static regime, it cannot be in the dynamic regime. A very simple argument (not reproduced here), assuming a non-absorbing material, is used to confirm this. He then puts forward a reformulation of the polarisability taking into account a phenomenon called the radiation reaction. This is the idea that, while an oscillating dipole certainly generates an electromagnetic field acting on neighbouring dipoles, it also generates an electric field which acts upon itself. A mathematical expression for this radiation reaction electric field is obtained by the following arguments. To begin with we recall that a moving charged particle, with charge q and position vector $\mathbf{r}(t)$ emits radiation into the surrounding medium (labelled 2), with instantaneous power given by the Larmor formula [4]:

$$\Psi(t) = \frac{q^2}{6\pi\varepsilon_2 c_2^3}\left(\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right)^2 . \tag{7.97}$$

Since the particle emits power by radiation, it must be losing mechanical energy. This argument leads one naturally to introduce an instantaneous force $\mathbf{F}_{\mathrm{rad}}(t)$ called the radiation reaction force into the equation of motion of the particle. It remains only to determine this radiation reaction force, by integrating the power balance equation $\mathbf{F}_{\mathrm{rad}}\cdot\mathrm{d}\mathbf{r}/\mathrm{d}t = -\Psi$ between two times [4], which leads to the expression

$$\mathbf{F}_{\mathrm{rad}}(t) = \frac{q^2}{6\pi\varepsilon_2 c_2^3}\frac{\mathrm{d}^3\mathbf{r}}{\mathrm{d}t^3} . \tag{7.98}$$

This result takes on a remarkable form when the particle has an oscillating motion on a straight line. Imposing a motion of type $\mathbf{r}(t) = \mathbf{r}_0\exp(-\mathrm{i}\omega t)$ on the position vector $\mathbf{r}(t)$, the expression for $\mathbf{F}_{\mathrm{rad}}(t)$ becomes

$$\mathbf{F}_{\mathrm{rad}}(t) = \frac{q^2}{6\pi\varepsilon_2 c_2^3}(-\mathrm{i}\omega)^3\mathbf{r}(t) . \tag{7.99}$$

Introducing the oscillating dipole moment $\mathbf{P}(t) = q\mathbf{r}(t)$ constituted by the oscillating charged particle and the oscillating electric field $\mathbf{E}_{\mathrm{rad}}(t) = \mathbf{F}_{\mathrm{rad}}(t)/q$ producing the radiation reaction force $\mathbf{F}_{\mathrm{rad}}(t)$ on the particle of charge q, and using the notation

$k_2 = \omega/c_2$ for the wave vector in medium 2, we find the following expression for the oscillating field $\mathbf{E}_{\rm rad}(t)$:

$$\boxed{\mathbf{E}_{\rm rad}(t) = \frac{\mathrm{i}k_2^3}{6\pi\varepsilon_2}\mathbf{P}(t)} \tag{7.100}$$

In [19], Draine extends this result by considering that the field $\mathbf{E}_{\rm rad}(t)$ expressed above is also the one produced by a polarisable material volume element with oscillating dipole moment $\mathbf{P}(t)$ on itself. Then the relation $\mathbf{P}_j(t) = \alpha_j \mathbf{E}_{\rm loc}(\mathbf{r}_j,t)$, the starting point for obtaining the DDA equation (7.65) on p. 179, is replaced by

$$\mathbf{P}_j(t) = \alpha_j\Big[\mathbf{E}_{\rm loc}(\mathbf{r}_j,t) + \mathbf{E}_{\rm rad}(\mathbf{r}_j,t)\Big] = \alpha_j \mathbf{E}_{\rm loc}(\mathbf{r}_j,t) + \frac{\mathrm{i}k_2^3\alpha_j}{6\pi\varepsilon_2}\mathbf{P}_j(t)\,, \tag{7.101}$$

and this new relation immediately implies the definition of a new polarisability $\alpha_j^{\rm RR}$ that includes the effects of the radiation reaction:

$$\boxed{\mathbf{P}_j(t) = \alpha_j^{\rm RR}\mathbf{E}_{\rm loc}(\mathbf{r}_j,t)\,, \quad \alpha_j^{\rm RR} = \frac{\alpha_j}{1 - \mathrm{i}k_2^3\alpha_j/6\pi\varepsilon_2}} \tag{7.102}$$

In his calculations, Draine applies the above formulation to the Clausius–Mossotti polarisability

$$\alpha_j^{\rm CM} = 3\varepsilon_2 v_j \frac{\varepsilon_j - \varepsilon_2}{\varepsilon_j + 2\varepsilon_2}\,.$$

He thus replaces α_j by $\alpha_j^{\rm CM}$ in (7.102), and thereby obtains a refined polarisability $\alpha_j^{\rm CM-RR}$.

The Doyle Formulation

In his paper [24], the author notes that the expression for the Clausius–Mossotti polarisability resembles the coefficient a_1 in the Mie sequence when the quantities x and $|m|x$ are both very small compared to unity. Indeed (see Sect. 7.3.3),

$$x \ll 1\,, \quad |m|x \ll 1 \quad\Longrightarrow\quad a_1 \sim -\frac{2\mathrm{i}x^3}{3}\frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2}\,. \tag{7.103}$$

Hence the proposal by Doyle to replace $(\varepsilon_1 - \varepsilon_2)/(\varepsilon_1 + 2\varepsilon_2)$ by $3\mathrm{i}a_1/2x^3$ in the Clausius–Mossotti expression. The Clausius–Mossotti polarisability $\alpha^{\rm CM}$ is then replaced by a new polarisability $\alpha^{\rm A1}$, only valid in the case of a spherical particle, and calculated as follows:

$$\boxed{\alpha^{\rm A1} = 3\varepsilon_2 v\frac{3\mathrm{i}a_1}{2x^3} = \frac{6\mathrm{i}\pi\varepsilon_2}{k_2^3}a_1 = \frac{6\mathrm{i}\pi\varepsilon_2}{k_2^3}\frac{m\psi_1(mk_2a)\psi_1'(k_2a) - \psi_1(k_2a)\psi_1'(mk_2a)}{m\psi_1(mk_2a)\xi_1'(k_2a) - \xi_1(k_2a)\psi_1'(mk_2a)}} \tag{7.104}$$

where the volume v has been replaced by $4\pi a^3/3$, with a the radius of the spherical particle, and the size parameter x has been replaced by $k_2 a$. The advantage with this new formulation has been demonstrated by H. Okamoto [25]. In fact it extends the range of validity of DDA calculations to particles with larger dimensions, so that one may loosen the constraint mentioned on p. 178 [see the criterion (7.60)] regarding the level of spatial discretisation. Moreover, C.E. Dungey and C.F. Bohren [26] have pointed out that the expression (7.104) for the polarisability α^{A1} includes a correction for radiation reaction effects identical to the one proposed by Draine.

7.4.3 Applying the Discrete Dipole Approximation

Now that we know the basic idea of the DDA, let us discuss some results of calculations based upon it. There are two parts here. The first aims to validate the DDA technique by confrontation with the Mie theory, while the second will illustrate the way the DDA can be used to model the radiative properties of nanoporous silica matrices. The reader interested in either of these two issues will find much more information in [22].

Spatial Discretisation of Spherical Particles and Comparison Between the DDA and Mie Theory

Since Mie theory applies only to spherical particles, we shall present here the results of DDA calculations applied to material balls discretised into an ensemble of small, polarisable volume elements, assuming the host medium to be the vacuum. Having completed the DDA calculations, we will be able to evaluate the radiative response of our material balls, then compare this data with the results of Mie calculations.

For the first series of calculations, the material ball is divided up into a relatively small number N of volume elements, in fact only 365. These elements are cubes in contact with one another, with side a and centres located at coordinates (ai, aj, ak), where i, j, and k are integers satisfying $i^2 + j^2 + k^2 < 20$. Figure 7.7 shows a 3D representation of the discretised material ball.

Having carried out the spatial discretisation, the following input data are needed to conduct the DDA calculation:

- the side a of each cubic volume element,
- the wavelength λ,
- the complex optical index $m = n + \mathrm{i}\chi$ of the material making up the ball at this wavelength,
- the unit vector $\mathbf{u}$ indicating the direction of the incident radiation,
- the model chosen to express the polarisability α.

For all calculations to be described in the following, the wavelength λ is fixed at the value 30 μm, the complex optical index m at $1.12 + \mathrm{i}0.017$ (these are the values

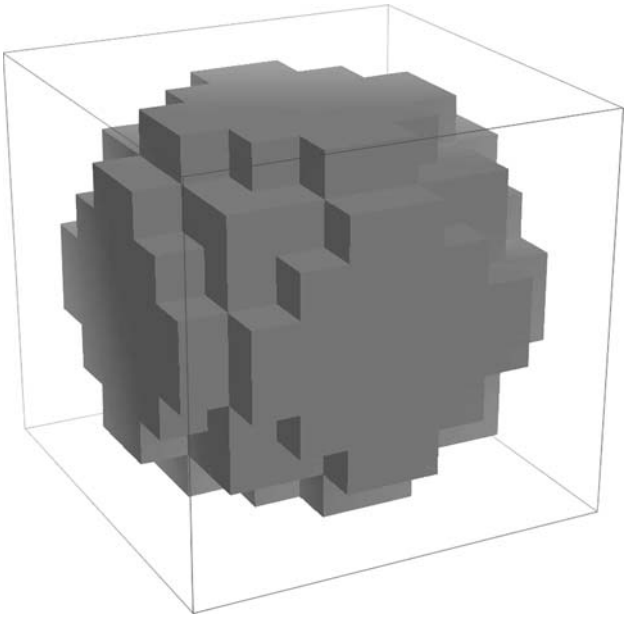

Fig. 7.7 Three-dimensional material ball divided into 365 discrete cubic volume elements

for sodium chloride at $\lambda = 30$ μm), and the polarisability is modelled by the Clausius–Mossotti formula, but taking into account radiation reaction effects, i.e., we use the $\alpha^{\mathrm{CM-RR}}$ formulation. (Note in passing that, strictly speaking, this model $\alpha^{\mathrm{CM-RR}}$ is not applicable here, because the volume elements of the spatial discretisation are not spherical.) The orthonormal basis $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ required for the DDA calculation of the cross-sections $C^{\mathrm{NP}}_{\mathrm{abs}}(\mathbf{u}, \mathrm{DDA})$ [see (7.91)], $C^{\mathrm{NP}}_{\mathrm{sca}}(\mathbf{u}, \mathrm{DDA})$ [see (7.79)], and $C^{\mathrm{NP}}_{\mathrm{ext}}(\mathbf{u}, \mathrm{DDA})$ [see (7.75)], and the phase function $\boldsymbol{\Phi}^{\mathrm{NP}}(\mathbf{u}, \mathbf{n}, \mathrm{DDA})$ [see (7.83)] is obtained from the orthonormal basis of the coordinate frame $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ by two successive rotations. The first, denoted by R_ζ, is through an angle $\zeta \in [0, 2\pi]$ about $\mathbf{x}$, while the second, denoted by R_ξ, is through an angle $\xi \in [0, \pi]$ about $R_\zeta(\mathbf{z})$. This means that the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ have the following components relative to the basis $(\mathbf{x}, \mathbf{y}, \mathbf{z})$:

$$\mathbf{u} = \begin{pmatrix} \cos\xi \\ \sin\xi\cos\zeta \\ \sin\xi\sin\zeta \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} -\sin\xi \\ \cos\xi\cos\zeta \\ \cos\xi\sin\zeta \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} 0 \\ -\sin\zeta \\ \cos\zeta \end{pmatrix}, \tag{7.105}$$

and the angles ξ and ζ are fixed at 1 and 2 radians, respectively, in our calculations. It remains only to specify the numerical values chosen for the side a. We vary this parameter between 0.1 and 10 μm. Under these conditions, Draine's value for $|m_1|k_0 a$ with $m_1 = m$ and $k_0 = 2\pi/\lambda$, as discussed on p. 177 ff, will vary between 0.023 and 2.3, and the criterion (7.60) for the validity of our DDA calculations,

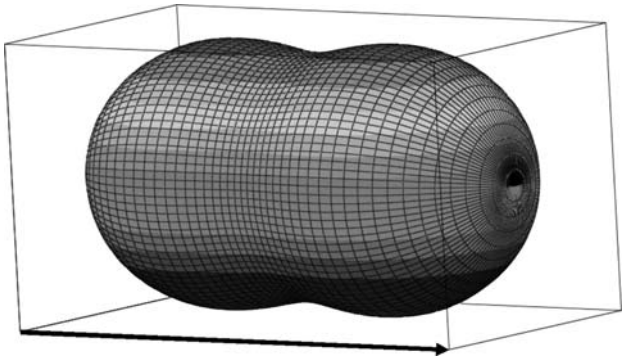

Fig. 7.8 Phase function $\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n},\mathrm{DDA})$ generated by the DDA for the spatial discretisation in Fig. 7.7. The side of the cubes in this discretisation is $a = 0.1$ μm. The *black arrow* shows the direction $\mathbf{u}$ of illumination

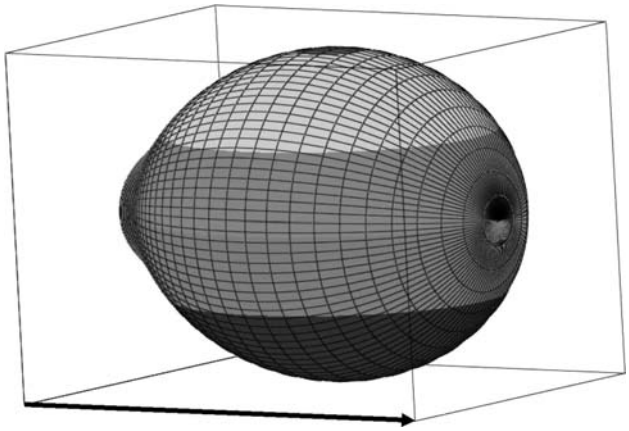

Fig. 7.9 As in Fig. 7.8, but with $a = 1.5$ μm

comfortably satisfied for small values of a, will be much less so when the side a has values of the order of a few μm.

The material ball containing the same volume of matter as the ensemble of our N cubic dipoles is characterised by a radius A defined by

$$\frac{4\pi A^3}{3} = Nv = Na^3\,, \qquad \text{whence} \quad A = a\sqrt[3]{\frac{3N}{4\pi}}\,. \tag{7.106}$$

The Mie calculations are carried out on this ball of radius A, with size parameter X given by

$$X = \frac{2\pi A}{\lambda} = \frac{2\pi a}{\lambda}\sqrt[3]{\frac{3N}{4\pi}}\,, \tag{7.107}$$

and varying between 0.092 and 9.2 when a varies between 0.1 and 10 μm.

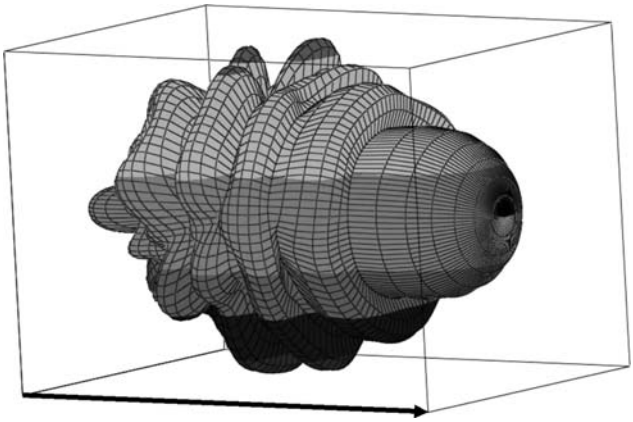

Fig. 7.10 As in Fig. 7.8, but with $a = 10$ μm. For better visibility, the surface plotted here is scanned, not by the radial vector $\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n},\mathrm{DDA})\mathbf{n}$, but by the radial vector $\log\left[10^6\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n},\mathrm{DDA})\right]\mathbf{n}$

Figures 7.8, 7.9, and 7.10 are direction indicators for the scattering phase functions $\Phi^{\mathrm{NP}}(\mathbf{u},\mathbf{n},\mathrm{DDA})$ generated by the DDA for the three values 0.1, 1.5, and 10 μm of the side a. Considering these three surfaces, the main points to note are as follows:

- The bigger a, the more directional the scattering becomes in the forward direction. Indeed, the surface in Fig. 7.10 corresponding to $a = 10$ μm had to be plotted on a logarithmic scale, otherwise only an extremely dominant lobe in the illumination direction $\mathbf{u}$ would have been visible. This agrees perfectly with the conclusions drawn from the numerical studies of the Mie theory in Sect. 7.3.6 (see in particular Fig. 7.6).
- When the value of a is not too high (see Figs. 7.8 and 7.9), the phase functions predicted by the DDA agree perfectly with those predicted by the Mie theory, and they have cylindrical symmetry about the axis $(\mathrm{O},\mathbf{u})$. The probability of scattering in a given direction $\mathbf{n}$ only depends on the angle θ between this direction $\mathbf{n}$ and the direction of illumination $\mathbf{u}$. This is no longer the case when the side a starts to get big (see Fig. 7.10), and with reason: when a is no longer negligible compared with the wavelength λ, the incident illumination is able to 'read', or 'resolve', the spatial discretisation it encounters, and it 'notices' that the discretised material volume with which it interacts in no way resembles a spherical particle.

Figure 7.11 shows the dependence of the ratio

$$\frac{C_{\mathrm{abs}}^{\mathrm{NP}}(\mathbf{u},\mathrm{DDA}) + C_{\mathrm{sca}}^{\mathrm{NP}}(\mathbf{u},\mathrm{DDA})}{C_{\mathrm{ext}}^{\mathrm{NP}}(\mathbf{u},\mathrm{DDA})}$$

on the side a. Values of this ratio very close to unity and up to parameter values $a = 3$ μm provide evidence for the consistency of the three formulas (7.75), (7.79), and (7.91) predicted by the DDA. In contrast, for large values of a, the three

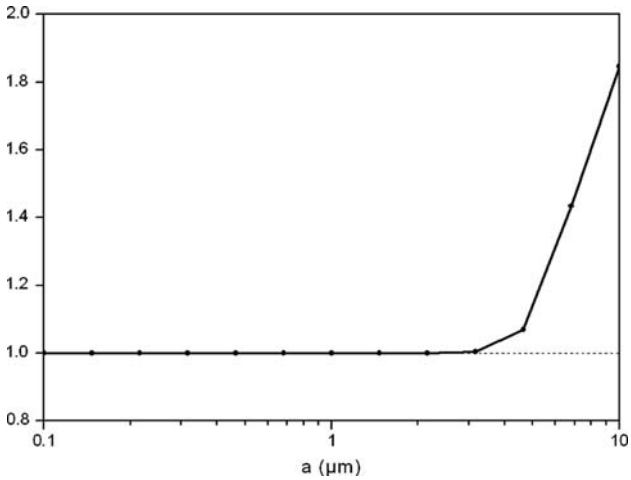

Fig. 7.11 Dependence of the ratio $\left[C_{\text{abs}}^{\text{NP}}(\mathbf{u},\text{DDA}) + C_{\text{sca}}^{\text{NP}}(\mathbf{u},\text{DDA})\right]/C_{\text{ext}}^{\text{NP}}(\mathbf{u},\text{DDA})$ on the side a

cross-sections resulting from the DDA calculations no longer correspond. This disagreement may be a consequence of the level reached by Draine's product $|m_1|k_0 a$ (0.70 when a is 3 μm), but it may also find an explanation in the difficulty in numerical evaluation of the directional integral appearing in (7.79) for the scattering cross-section $C_{\text{sca}}^{\text{NP}}(\mathbf{u},\text{DDA})$ when the scattering phenomenon has a marked directionality. (In the framework of the DDA calculations presented here, we have used the S_8 quadrature with 10 discrete ordinates per octant for numerical evaluation of the directional integrals. It may be that the use of angle quadratures specifically adapted to the strong forward scattering regime – see, for example, the work presented in [27] – would lead to improved consistency between the various cross-sections output by the DDA calculations for large values of the side a.)

If on the basis of the above discussion we choose to evaluate the scattering cross-section $C_{\text{sca}}^{\text{NP}}(\mathbf{u},\text{DDA})$, not using (7.79), but rather by the simple difference $C_{\text{ext}}^{\text{NP}}(\mathbf{u},\text{DDA}) - C_{\text{abs}}^{\text{NP}}(\mathbf{u},\text{DDA})$, then the three cross-sections output by the DDA compare rather well with their three Mie counterparts, as can be seen from Fig. 7.12, which shows on the same graph the dependence of the ratios

$$\frac{C_{\text{abs}}^{\text{NP}}(\mathbf{u},\text{DDA})}{C_{\text{abs}}^{\text{NP}}(\text{Mie})}, \qquad \frac{C_{\text{sca}}^{\text{NP}}(\mathbf{u},\text{DDA})}{C_{\text{sca}}^{\text{NP}}(\text{Mie})}, \qquad \frac{C_{\text{ext}}^{\text{NP}}(\mathbf{u},\text{DDA})}{C_{\text{ext}}^{\text{NP}}(\text{Mie})},$$

on the side a (with the obvious notation). For values of a less than or equal to 3 μm, the disagreement between the two theories is less than 4%, but for large values of a, the two modelling techniques move further and further apart. There are two reasons for this, which we have already mentioned:

- When a is large, Draine's product $|m_1|k_0 a$ grows quite large, indicating that the spatial discretisation becomes rather approximate, and the applicability of the DDA can be reasonably brought into question.

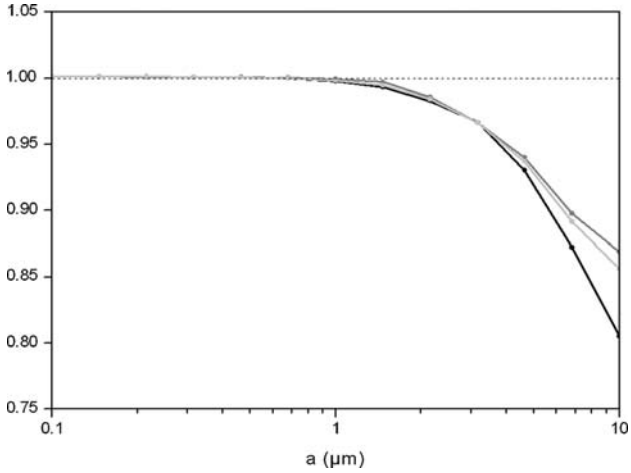

Fig. 7.12 Dependence of $C_{\mathrm{abs}}^{\mathrm{NP}}(\mathbf{u},\mathrm{DDA})/C_{\mathrm{abs}}^{\mathrm{NP}}(\mathrm{Mie})$ (*black curve*), $C_{\mathrm{sca}}^{\mathrm{NP}}(\mathbf{u},\mathrm{DDA})/C_{\mathrm{sca}}^{\mathrm{NP}}(\mathrm{Mie})$ (*dark grey curve*), and $C_{\mathrm{ext}}^{\mathrm{NP}}(\mathbf{u},\mathrm{DDA})/C_{\mathrm{ext}}^{\mathrm{NP}}(\mathrm{Mie})$ (*light grey curve*) on the side a

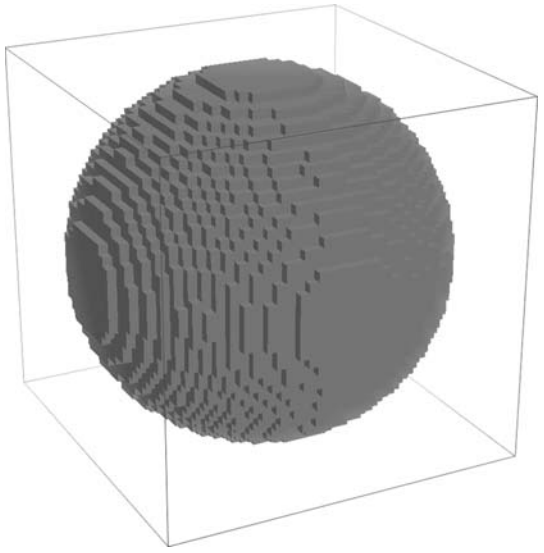

Fig. 7.13 Three-dimensional representation of the spatial discretisation refined to $N' = 33\,059$ cubic volume elements

- When a is no longer negligible compared with the wavelength λ, the spatial discretisation in Fig. 7.7 no longer looks like a spherical object from the point of view of the incident radiation.

The difficulties encountered with the DDA when the side a characterising the spatial discretisation in Fig. 7.7 is large are resolved by carrying out a finer spatial discretisation. (There is a price to pay for this in the significantly extended computation

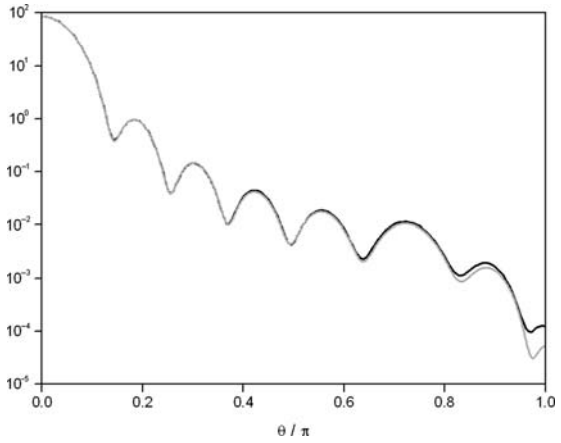

Fig. 7.14 *Black curve*: Phase function $\Phi^{\mathrm{NP}}(\theta, \mathrm{DDA})$ produced by the DDA for the spatial discretisation in Fig. 7.13. The side of the cubes in this discretisation is $a' = 2.23$ μm. *Grey curve*: Phase function $\Phi^{\mathrm{NP}}(\theta, \mathrm{Mie})$ produced by Mie theory applied to the equivalent material ball of radius $A = 44.3$ μm

time.) For example, in the delicate case where $a = 10$ μm, the material ball now being discretised into $N' = 33\,059$ cubic elements (see Fig. 7.13) of side a', where a' is defined so that the total volume of matter is conserved, i.e.,

$$N'a'^3 = Na^3 \,, \qquad \text{whence} \quad a' = a\sqrt[3]{\frac{N}{N'}} \approx 2.23 \text{ μm} \,,$$

the DDA calculation leads to the following excellent numerical results:

$$\frac{C_{\mathrm{abs}}^{\mathrm{NP}}(\mathbf{u}, \mathrm{DDA}) + C_{\mathrm{sca}}^{\mathrm{NP}}(\mathbf{u}, \mathrm{DDA})}{C_{\mathrm{ext}}^{\mathrm{NP}}(\mathbf{u}, \mathrm{DDA})} = \frac{C_{\mathrm{abs}}^{\mathrm{NP}}(\mathbf{u}, \mathrm{DDA})}{C_{\mathrm{abs}}^{\mathrm{NP}}(\mathrm{Mie})} = \frac{C_{\mathrm{sca}}^{\mathrm{NP}}(\mathbf{u}, \mathrm{DDA})}{C_{\mathrm{sca}}^{\mathrm{NP}}(\mathrm{Mie})} = \frac{C_{\mathrm{ext}}^{\mathrm{NP}}(\mathbf{u}, \mathrm{DDA})}{C_{\mathrm{ext}}^{\mathrm{NP}}(\mathrm{Mie})}$$
$$= 1.000 \,.$$

Regarding the phase function produced by this calculation, it agrees perfectly with the Mie calculation, as can be seen from the graph in Fig. 7.14.

Evaluating the Radiative Response of Nanoporous Silica Matrices

Many studies of nanoporous thermal superinsulating materials have been carried out at CEA/Le Ripault (France) due to their extraordinarily low level of effective thermal conductivity, namely, a few $\mathrm{mW\,m^{-1}K^{-1}}$ at room temperature and under primary air vacuum. The first step in making these highly porous materials (their porosity, exclusively open, is of the order of 90%) is the production of a nanoporous matrix by compacting fumed silica nanoparticles with diameters of the order of 10 nm. The nanoporous matrix is the majority solid component of this family

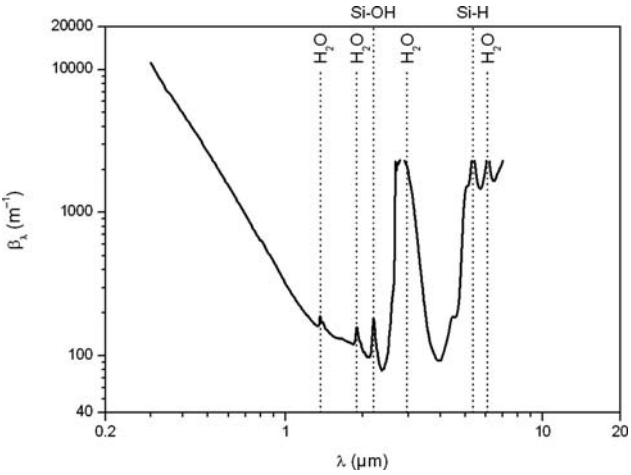

Fig. 7.15 Spectrum β_λ of the extinction coefficient, evaluated for a sample of thickness 2.0 mm made from the commercially available powder Wacker HDK-T30

of materials, with a mass fraction over 80%. Very small volume fractions (of the order of one percent) of microscale components are dispersed within the matrix to enhance its infrared opacity and mechanical strength. (A detailed microstructural description and TEM images of a typical nanoporous superinsulator can be found in [28].) Since the heat transfer by conduction has been minimised in this family of materials (in particular by judicious choice of the type, size, and arrangement of the various constituents that make them up), this makes it important to determine the level of heat transfer by radiation within them. Indeed, since the conduction transport is highly attenuated, it may turn out to be superseded by radiative transport, and this even at relatively low temperatures, e.g., room temperature.

In order to understand the radiative transfer properties within nanoporous thermal superinsulators, we decided to simplify the materials by restricting to the nanoporous matrix, i.e., without the microscale additives, and to quantify the bulk radiative properties of these matrices by suitable experiments and models.

The nanoporous matrix samples were produced from various commercially available fumed silica powders. These samples were then subjected to hemispherical-directional transmission and reflection spectrophotometric analysis over the spectral band 0.2–20 μm. Then, inverting the radiative transfer equation (see the discussion of this method in [29]), we deduced from the experimental data R_λ and T_λ the bulk radiative properties of the matrices, namely, the spectra β_λ and ω_λ of their extinction coefficient and scattering albedo. As an illustration, the graph in Fig. 7.15 shows the spectrum β_λ evaluated for a sample of thickness 2.0 mm made from the powder HDK-T30 commercialised by the German company Wacker.

Having done this, we wanted to find out whether the experimental spectra β_λ and ω_λ would have been adequately predicted by Mie theory. For all the nanoporous

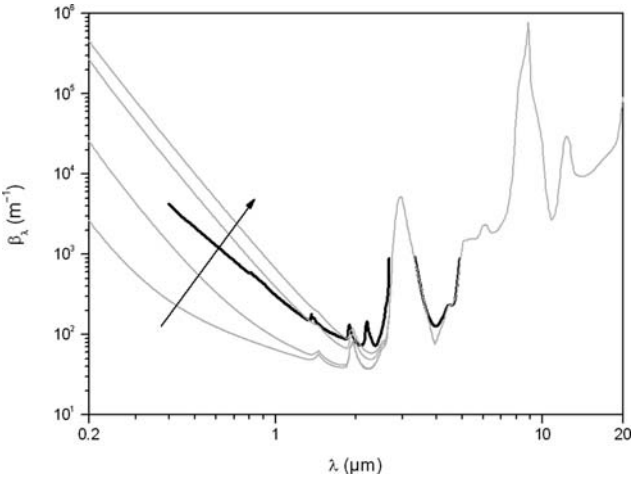

Fig. 7.16 *Black curve*: Experimental spectrum β_λ obtained for a sample of the silica powder Wacker HDK-T30 of thickness 5.0 mm. *Grey curves*: Spectra β_λ produced by Mie calculations for scatterer diameters 9, 20, 45, and 55 nm. The *arrow* indicates increasing values of the diameter

matrices we investigated, the comparison between theory and experiment led to the following three observations:

- The Mie spectra β_λ always agree very well qualitatively and quantitatively with their experimental counterparts for long wavelengths (typically $\lambda \geq 2$ μm), and this whatever the diameter (from ten to a hundred or so nm) chosen for the spherical scatterer in the Mie calculations, provided that one respects the experimental data constituted by the solid volume fraction in the sample in these calculations.
- The Mie spectra β_λ and ω_λ can be matched qualitatively with their experimental counterparts at medium wavelengths, i.e., in the μm range, provided that the spherical Mie scatterer is attributed a significantly larger diameter (of the order of 5 times greater) than the silica nanoparticles making up the samples.
- At short wavelengths (typically $\lambda \leq 1$ μm), even with the optimal scatterer diameter in the Mie calculations, this theory is unable to account adequately for the dependence of the extinction coefficient β_λ on the wavelength. In every case, the spectrum β_λ predicted by the Mie theory over this spectral range decreases much faster than its experimental counterpart.

These three observations are described in much more detail in [22]. The graphs in Figs. 7.16 and 7.17 are intended to clarify these points. These two graphs, β_λ in Fig. 7.16 and ω_λ in Fig. 7.17, are the experimental spectra obtained on a sample of the silica powder Wacker HDK-T30 of thickness 5.0 mm together with the spectra produced by Mie calculations for scatterer diameters between 9 nm (average diameter of the primary nanoparticle in the powder HDK-T30 estimated from the value of the specific surface area of this powder) and 55 nm. One can make the following observations concerning these two figures:

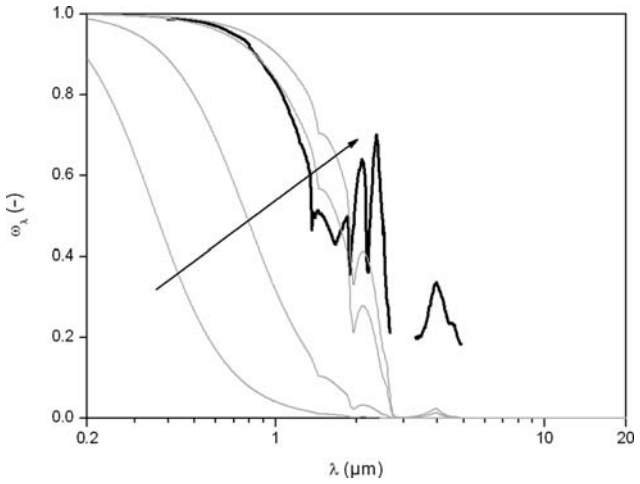

Fig. 7.17 *Black curve*: Experimental spectrum ω_λ obtained for a sample of the silica powder Wacker HDK-T30 of thickness 5.0 mm. *Grey curves*: Spectra ω_λ produced by Mie calculations for scatterer diameters 9, 20, 45, and 55 nm. The *arrow* indicates increasing values of the diameter

- The Mie spectra β_λ and ω_λ do not agree at all with their experimental counterparts at wavelengths $\lambda \leq 2$ μm when the diameter of the Mie scatterer is taken as 9 nm.
- The experimental spectrum ω_λ exhibits a clear transition from a dominant scattering regime at short wavelengths to a dominant absorption regime at long wavelengths. The spectral position of this transition justifies the definition of an effective Mie diameter D_ω, estimated here at 45 nm.
- The Mie spectrum β_λ can be fairly well matched to its experimental counterpart at wavelengths in the μm range for an effective Mie diameter D_β of 55 nm, a value in good agreement with the value D_ω resulting from analysis of the experimental spectrum ω_λ.
- Even if we attribute one of the optimal values D_ω and D_β introduced above to the diameter D of the Mie scatterer, the Mie theory is unable to account for the slowly decreasing experimental spectrum β_λ at wavelengths $\lambda \leq 1$ μm.

These observations lead us to the following conclusions:

- At wavelengths greater than 1 μm, the Mie spectra β_λ and ω_λ agree adequately with their experimental counterparts for a value D of the Mie scatterer diameter taken for example as $(D_\omega + D_\beta)/2$. This implies that, for wavelengths greater than 1 μm, our nanoporous matrices behave, from the point of view of the matter–radiation interaction, exactly like populations of spherical particles. The effective diameters D of these particles, resulting from comparison between the Mie theory and experiment, are significantly greater than those of the nanoparticles and give rise to the idea of a representative cluster, defined as a cluster of nanoparticles containing the same volume of matter as a ball of diameter D.

- When the wavelength is large compared with the size of the fictional box enclosing the representative cluster, the matter–radiation interaction is barely sensitive to the size, shape, or arrangement of the matter within the cluster. This is because the latter appears to be pointlike as far as the wavelength of the radiation is concerned.
- However, this is no longer the case when the wavelength has the same order of magnitude as the size of the box enclosing the cluster. In this case, theoretical prediction of the radiative response of nanoporous matrices can no longer go through the Mie theory, but must use a modelling technique sensitive to the spatial distribution of matter within the representative cluster.

To check this last claim, we began to develop computer programs (i) for generating representative nanoparticle clusters and (ii) for calculating the radiative properties of these clusters using the DDA.

For the computer generation of representative clusters, we took into account data concerning the optimal diameter D deduced from the above Mie analyses, but also 'global' information available to us concerning our nanoporous matrix samples, viz., their density (around 300 kg/m^3) and fractal dimensions (1.8 for the nanometric fumed silica structures, according to the literature [30]). Since the representative cluster contains by definition the same volume of matter as a ball of diameter D, knowledge of this parameter can be used to calculate the number N of nanoparticles making up the representative cluster via the relation $N = (D/d)^3$, where d is the diameter of the nanoparticles. Furthermore, the occupation volume V associated with the representative cluster follows directly from the density ρ_m of the matrix via the formula

$$V = N\frac{\pi d^3}{6}\frac{\rho_\mathrm{np}}{\rho_\mathrm{m}}\,,$$

where ρ_np is the density of the material making up the nanoparticles.

Since the number of nanoparticles N and the occupation volume V of the representative cluster are known, it can be computer generated by means of two standard algorithms for particle aggregation commonly called DLA and DLCCA, which stand for diffusion-limited aggregation and diffusion-limited cluster–cluster aggregation, respectively. The idea behind these two algorithms is to allow the N nanoparticles to diffuse, i.e., move randomly, within a fictional cubic box in a way governed by rules specific to each algorithm. To begin with, we used the DLCCA algorithm to generate our representative clusters, because this algorithm is known to produce fractal structures with fractal dimension close to 1.8 in a 3D space. However, our representative clusters being very small, generally comprising two or three hundred nanoparticles at the most, we observed that their fractal structure was far from clear, so that it was extremely difficult, even impossible, to evaluate their fractal dimension. We subsequently used other aggregation algorithms, and in particular the DLA algorithm, to generate our representative clusters.

As an illustration, we found above for the Wacker HDK-T30 silica sample of thickness 5.0 mm and density $\rho_\mathrm{m} = 290\ \mathrm{kg/m^3}$ a diameter D equal to 45 nm or 55 nm, depending on the choice of experimental spectrum (extinction coefficient β_λ or

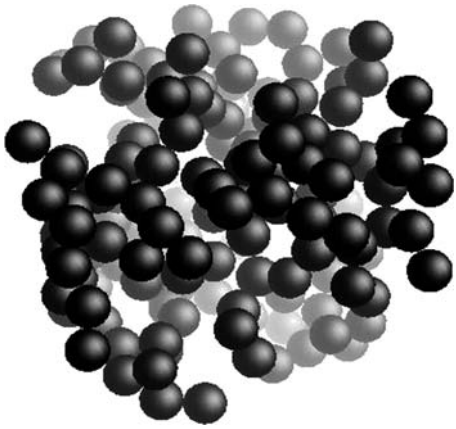

Fig. 7.18 Representative cluster of the Wacker HDK-T30 silica powder sample with thickness 5.0 mm, generated using the DLCCA algorithm

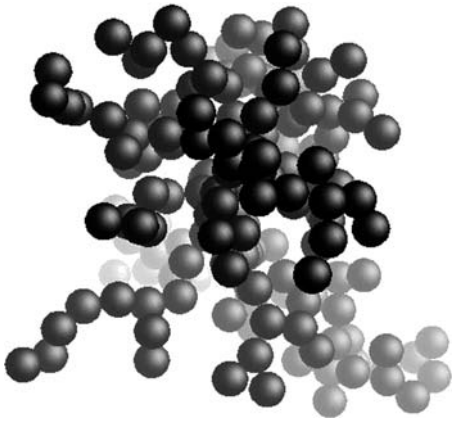

Fig. 7.19 Representative cluster of the Wacker HDK-T30 silica powder sample with thickness 5.0 mm, generated using the DLA algorithm

scattering albedo ω_λ) exploited via the Mie theory. Selecting the average $D = 50$ nm, and recalling that, for the Wacker HDK-T30 powder, the diameter d of the nanoparticles is about 9 nm, the calculations lead to a number $N = 171$ of nanoparticles in the representative cluster and an occupation volume $V = 496\,500$ nm^3 associated with this cluster, implying a cubic box of side 79 nm. On the basis of this data, the DLCCA and DLA algorithms produced clusters like those shown in Figs. 7.18 and 7.19.

Once the representative cluster had been generated, we subjected this material structure to DDA calculations covering our chosen spectral range 0.2–20 μm. To do this, each nanoparticle in the cluster was treated as a single oscillating

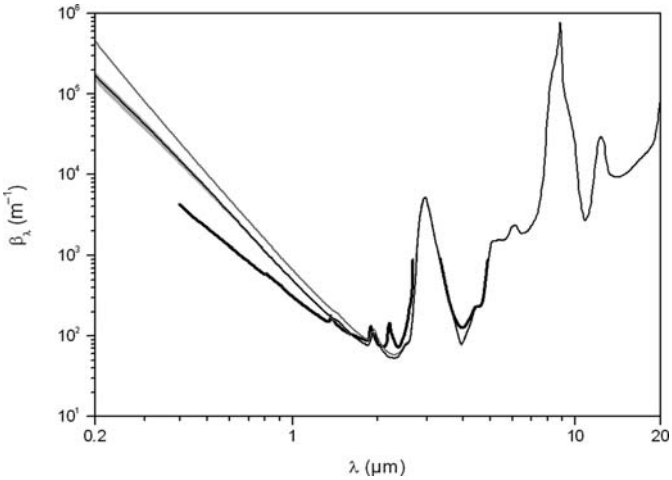

Fig. 7.20 Spectra of the extinction coefficient β_λ associated with the Wacker HDK-T30 silica sample of thickness 5.0 mm. *Thick black curve*: Spectrum derived from optical measurements carried out on the sample. *Dark grey curve*: Spectrum produced by a Mie calculation with a diameter D for the Mie scatterer fixed at 55 nm. *Light grey band* and *thin black curve*: Results of a DDA calculation carried out on the DLCCA cluster of Fig. 7.18

electromagnetic dipole. This assumption is justified here by the fact that, since the nanoparticles in our powders are very small, with diameters $d \sim 10$ nm, the Draine product $|m|kd$ of p. 177 ff is always smaller than unity, and the criterion (7.60) for the validity of the DDA calculation is thus satisfied. For example, for the Wacker HDK-T30 silica powder, the maximal value of the product $|m|kd$ over the range 0.2–20 μm is only 0.44 (reached when $\lambda = 0.2$ μm).

For each wavelength λ and for a whole range of previously defined illumination directions $\mathbf{u}$ covering 2π steradians, the extinction cross-section $C_{\text{ext}}^{\text{NP}}(\mathbf{u}, \lambda, \text{DDA})$ and scattering cross-section $C_{\text{sca}}^{\text{NP}}(\mathbf{u}, \lambda, \text{DDA})$ of the representative cluster under non-polarised illumination were calculated using the DDA. The directional monochromatic extinction coefficient and scattering albedo $\beta_\lambda(\mathbf{u})$ and $\omega_\lambda(\mathbf{u})$ were then deduced from these results using the relations

$$\beta_\lambda(\mathbf{u}) = \frac{C_{\text{ext}}^{\text{NP}}(\mathbf{u}, \lambda, \text{DDA})}{V}$$

and

$$\omega_\lambda(\mathbf{u}) = \frac{C_{\text{sca}}^{\text{NP}}(\mathbf{u}, \lambda, \text{DDA})}{C_{\text{ext}}^{\text{NP}}(\mathbf{u}, \lambda, \text{DDA})} ,$$

where V is the occupation volume associated with the representative cluster.

The results of these calculations are shown in Figs. 7.20 and 7.21, together with the experimental spectrum β_λ obtained for the Wacker HDK-T30 silica sample of thickness 5.0 mm and the best spectrum β_λ predicted by Mie theory (corresponding to the diameter $D = 55$ nm for the Mie scatterer). Superposed on these two spectra

are the results of the DDA calculations applied to the DLCCA representative cluster of Fig. 7.18 for the graph of Fig. 7.20 and the DLA representative cluster of Fig. 7.19 for the graph of Fig. 7.21. At long wavelengths, the DDA and Mie theory produce absolutely identical extinction coefficients, which is perfectly consistent with the above assertion concerning the pointlike appearance of the representative cluster at the wavelength of the radiation. However, when the wavelength is short enough to begin to probe the spatial arrangement of the matter within the representative cluster, the DDA predicts significantly smaller extinction coefficients than Mie theory, and approaches the experimental values.

The graphs of Figs. 7.20 and 7.21 illustrate the dispersion in the values $\beta_\lambda(\mathbf{u})$ predicted by the DDA for different values of the illumination direction $\mathbf{u}$. These dispersions show up as bands of values β_λ of varying widths containing the average curves defined arbitrarily as arithmetic means of the $\beta_\lambda(\mathbf{u})$ over all the directions $\mathbf{u}$ examined. Here it is interesting to note that the dispersion in the values $\beta_\lambda(\mathbf{u})$ is much smaller with the DLCCA cluster than with the DLA cluster. This point is related to the rather 'isotropic' visual appearance of the DLCCA cluster, whereas the DLA cluster seems to be built up from particle branches, picking out preferred directions and inevitably producing larger variations in its properties for different directions of observation.

The DDA calculations presented here clearly demonstrate the effect of the spatial arrangement of the nanoparticles on the radiative properties of nanoporous matrices. This modelling technique thus proves to be very sensitive to the way in which the nanoparticles are arranged within the representative cluster. If they were concentrated within a single nucleus, the DDA would generate the Mie spectrum for β_λ, whereas when they are distributed according to the two different aggregation algorithms, the DDA generates significantly different spectra β_λ.

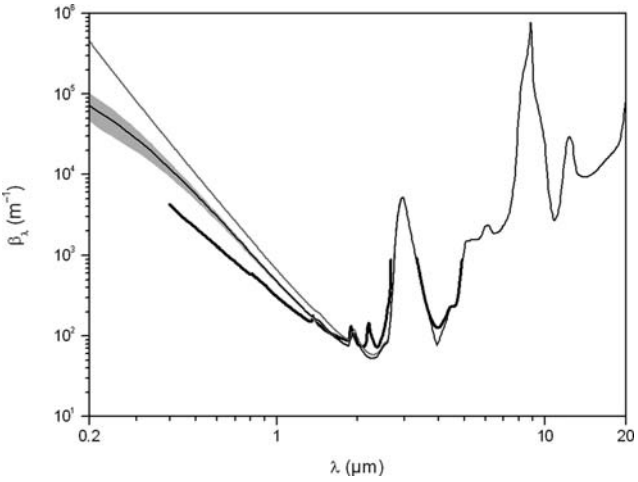

Fig. 7.21 As in Fig. 7.20, but the DDA calculations were carried out on the DLA cluster of Fig. 7.19

7.5 Summary

In this chapter, we have begun to outline the general theory of the interaction between electromagnetic radiation and a material particle of arbitrary shape. This outline provides the tools needed to solve the Mie electromagnetic problem in the case of spherical particles. And through a specific example, we have seen how to apply the Mie theory to calculate the radiative response of populations of spheres distributed randomly with a small enough volume fraction in space. Finally, we have discussed the interaction between electromagnetic radiation and a spatially structured material aggregate. Since the Mie theory fails in this type of situation, we have described an alternative modelling technique called the discrete dipole approximation, which can handle such interaction configurations. The advantages of this new theory over the Mie model have been exemplified by the study of nanoporous silica matrices, when the wavelength of the incident radiation is small enough to probe the spatial arrangement of the matter within the matrix.

Appendix: Analytical Solution of Mie's Electromagnetic Problem

The solution is rather tedious, and suitable accounts can be found in [3, 8]. In this appendix, we shall merely outline the four main steps of the solution. Note that, throughout this discussion, we shall simplify the notation by omitting the temporal harmonic term $\exp(-\mathrm{i}\omega t)$.

First Stage: General Solution of a Scalar Propagation Equation in Spherical Coordinates

To begin with, consider a spatial scalar field $\psi(\mathbf{r})$ satisfying the propagation equation

$$\nabla^2\psi + k^2\psi = 0\,, \tag{7.108}$$

in a region of space with spherical symmetry. It is then natural to introduce the standard spherical coordinates (r,θ,φ), where $r \geq 0$, $0 \leq \theta \leq \pi$, and $0 \leq \varphi \leq 2\pi$, whence it can be shown that, in this coordinate system, the general solution of (7.108) has the form

$$\boxed{\psi(r,\theta,\varphi) = \sum_{m=0}^{\infty}\sum_{n=m}^{\infty}\Big[\text{linear combination of } \psi_{\mathrm{e}mn}(r,\theta,\varphi) \text{ and } \psi_{\mathrm{o}mn}(r,\theta,\varphi)\Big]} \tag{7.109}$$

where m ranges over $\mathbb{Z}^+$ and n ranges from m to ∞, and the two series of functions $\psi_{\mathrm{e}mn}(r,\theta,\varphi)$ and $\psi_{\mathrm{o}mn}(r,\theta,\varphi)$, called scalar spherical harmonics, are given by

$$\boxed{\begin{aligned}\psi_{\mathrm{e}mn}(r,\theta,\varphi) &= z_n(kr)P_n^m(\cos\theta)\cos(m\varphi)\,,\\ \psi_{\mathrm{o}mn}(r,\theta,\varphi) &= z_n(kr)P_n^m(\cos\theta)\sin(m\varphi)\,,\end{aligned}} \tag{7.110}$$

in which z_n is any linear combination of the two spherical Bessel functions j_n and y_n, and P_n^m are the associated Legendre functions of the first kind. The subscripts e and o on the harmonic functions $\psi_{\mathrm{e}mn}(r,\theta,\varphi)$ and $\psi_{\mathrm{o}mn}(r,\theta,\varphi)$, adopted from the account in [3], remind us that these are even and odd functions of the angle φ, respectively.

Second Stage: General Expression for an Electromagnetic Field in Spherical Coordinates

Still working in a spherically symmetric spatial region, we now consider an electromagnetic field $\{\mathbf{E}(\mathbf{r}),\mathbf{H}(\mathbf{r})\}$ satisfying the standard Maxwell equations

$$\nabla\cdot\mathbf{E}=0\,,\quad \nabla\cdot\mathbf{H}=0\,,\quad \nabla\times\mathbf{E}=\mathrm{i}\omega\mu\mathbf{H}\,,\quad \nabla\times\mathbf{H}=-\mathrm{i}\omega\varepsilon\mathbf{E}\,. \tag{7.111}$$

These imply the propagation equations

$$\nabla^2\mathbf{E}+k^2\mathbf{E}=\mathbf{0}\,,\qquad \nabla^2\mathbf{H}+k^2\mathbf{H}=\mathbf{0}\,,\qquad k=\omega\sqrt{\varepsilon\mu}\,. \tag{7.112}$$

It can be shown that the general solution for the electric field $\mathbf{E}(r,\theta,\varphi)$ can be expressed in terms of four sequences of vector functions $\mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi)$, $\mathbf{M}_{\mathrm{o}mn}(r,\theta,\varphi)$, $\mathbf{N}_{\mathrm{e}mn}(r,\theta,\varphi)$, and $\mathbf{N}_{\mathrm{o}mn}(r,\theta,\varphi)$, called vector spherical harmonics:

$$\boxed{\mathbf{E}(r,\theta,\varphi)=\sum_{m=0}^{\infty}\sum_{n=m}^{\infty}\left[\begin{array}{c}\text{linear combination of } \mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi),\ \mathbf{M}_{\mathrm{o}mn}(r,\theta,\varphi),\\ \mathbf{N}_{\mathrm{e}mn}(r,\theta,\varphi),\ \text{and } \mathbf{N}_{\mathrm{o}mn}(r,\theta,\varphi)\end{array}\right]} \tag{7.113}$$

These four sequences of vector harmonics are related to the two series of scalar harmonics (7.110). If $(\mathbf{e}_r,\mathbf{e}_\theta,\mathbf{e}_\varphi)$ is the local right-handed orthonormal basis associated with the spherical coordinates (r,θ,φ), then the vector spherical harmonics are specified by

$$\mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi)=\nabla\times\Big[\psi_{\mathrm{e}mn}(r,\theta,\varphi)\mathbf{e}_r\Big]\,,\quad \mathbf{M}_{\mathrm{o}mn}(r,\theta,\varphi)=\nabla\times\Big[\psi_{\mathrm{o}mn}(r,\theta,\varphi)\mathbf{e}_r\Big]\,,$$

whence

$$\boxed{\mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi)=-mz_n(kr)\frac{P_n^m(\cos\theta)}{\sin\theta}\sin(m\varphi)\mathbf{e}_\theta-z_n(kr)\frac{\mathrm{d}P_n^m(\cos\theta)}{\mathrm{d}\theta}\cos(m\varphi)\mathbf{e}_\varphi} \tag{7.114}$$

$$\boxed{\mathbf{M}_{omn}(r,\theta,\varphi) = mz_n(kr)\frac{P_n^m(\cos\theta)}{\sin\theta}\cos(m\varphi)\mathbf{e}_\theta - z_n(kr)\frac{\mathrm{d}P_n^m(\cos\theta)}{\mathrm{d}\theta}\sin(m\varphi)\mathbf{e}_\varphi} \tag{7.115}$$

and

$$\mathbf{N}_{emn}(r,\theta,\varphi) = \frac{\nabla\times\mathbf{M}_{emn}(r,\theta,\varphi)}{k}\,, \quad \mathbf{N}_{omn}(r,\theta,\varphi) = \frac{\nabla\times\mathbf{M}_{omn}(r,\theta,\varphi)}{k}\,,$$

whence

$$\boxed{\begin{aligned}\mathbf{N}_{emn}(r,\theta,\varphi) = {} & n(n+1)\frac{z_n(kr)}{kr}P_n^m(\cos\theta)\cos(m\varphi)\mathbf{e}_r \\ & + \frac{1}{kr}\frac{\mathrm{d}\left[rz_n(kr)\right]}{\mathrm{d}r}\frac{\mathrm{d}P_n^m(\cos\theta)}{\mathrm{d}\theta}\cos(m\varphi)\mathbf{e}_\theta \\ & - \frac{m}{kr}\frac{\mathrm{d}\left[rz_n(kr)\right]}{\mathrm{d}r}\frac{P_n^m(\cos\theta)}{\sin\theta}\sin(m\varphi)\mathbf{e}_\varphi\end{aligned}} \tag{7.116}$$

$$\boxed{\begin{aligned}\mathbf{N}_{omn}(r,\theta,\varphi) = {} & n(n+1)\frac{z_n(kr)}{kr}P_n^m(\cos\theta)\sin(m\varphi)\mathbf{e}_r \\ & + \frac{1}{kr}\frac{\mathrm{d}\left[rz_n(kr)\right]}{\mathrm{d}r}\frac{\mathrm{d}P_n^m(\cos\theta)}{\mathrm{d}\theta}\sin(m\varphi)\mathbf{e}_\theta \\ & + \frac{m}{kr}\frac{\mathrm{d}\left[rz_n(kr)\right]}{\mathrm{d}r}\frac{P_n^m(\cos\theta)}{\sin\theta}\cos(m\varphi)\mathbf{e}_\varphi\end{aligned}} \tag{7.117}$$

Finally, it can be shown that two vector spherical harmonics of type **M** and **N** with the same indices are related to one another via the relations

$$\boxed{\nabla\times\mathbf{M}(r,\theta,\varphi) = k\mathbf{N}(r,\theta,\varphi)\,, \quad \nabla\times\mathbf{N}(r,\theta,\varphi) = k\mathbf{M}(r,\theta,\varphi)} \tag{7.118}$$

Third Stage: Expansion of a Plane Wave in Vector Spherical Harmonics

We can now begin to consider the Mie problem, which concerns the interaction between some incident electromagnetic radiation $\left\{\mathbf{E}_{\mathrm{inc}}(\mathbf{r}),\mathbf{H}_{\mathrm{inc}}(\mathbf{r})\right\}$ and a spherical particle of radius a. The incident wave is assumed to be plane and polarised, with direction of propagation $\mathbf{u}$ and polarisation $\mathbf{e}$, where $\mathbf{e}\perp\mathbf{u}$. The incident electric field $\mathbf{E}_{\mathrm{inc}}(\mathbf{r})$ is then given by

$$\mathbf{E}_{\mathrm{inc}}(\mathbf{r}) = E_0\mathbf{e}\exp(\mathrm{i}k_2\mathbf{u}\cdot\mathbf{r})\,. \tag{7.119}$$

The two angles of the spherical coordinates (r,θ,φ) are defined as follows: θ is the angle $\theta = (\mathbf{u},\mathbf{r})$ between $\mathbf{u}$ and $\mathbf{r}$, with $0\le\theta\le\pi$, and φ is the polar angle with $\mathbf{e}$ of the orthogonal projection of $\mathbf{r}$ in the plane orthogonal to $\mathbf{u}$, with $0\le\varphi\le 2\pi$. In

this system of spherical coordinates,

$$\mathbf{E}_{\text{inc}}(\mathbf{r}) = E_0\big(\sin\theta\cos\varphi\,\mathbf{e}_r + \cos\theta\cos\varphi\,\mathbf{e}_\theta - \sin\varphi\,\mathbf{e}_\varphi\big)\exp(\mathrm{i}k_2 r\cos\theta)\ ,$$

and using the results of the last section, this field can be expressed as an expansion in terms of vector spherical harmonics:

$$\begin{aligned}\mathbf{E}_{\text{inc}}(r,\theta,\varphi) &= E_0\big(\sin\theta\cos\varphi\,\mathbf{e}_r + \cos\theta\cos\varphi\,\mathbf{e}_\theta - \sin\varphi\,\mathbf{e}_\varphi\big)\exp(\mathrm{i}k_2 r\cos\theta)\\ &= \sum_{m=0}^{\infty}\sum_{n=m}^{\infty}\Big[A_{\mathrm{e}mn}\mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi) + A_{\mathrm{o}mn}\mathbf{M}_{\mathrm{o}mn}(r,\theta,\varphi)\\ &\qquad + B_{\mathrm{e}mn}\mathbf{N}_{\mathrm{e}mn}(r,\theta,\varphi) + B_{\mathrm{o}mn}\mathbf{N}_{\mathrm{o}mn}(r,\theta,\varphi)\Big]\ .\end{aligned} \tag{7.120}$$

The functions $z_n(kr)$ appearing in the expressions for these harmonics must be taken equal to the functions $j_n(k_2 r)$ for the following reasons:

- the incident electromagnetic field propagates in the host medium (labelled 2),
- when $\rho \to 0$, $y_n(\rho) \to -\infty$, whence the functions y_n must be rejected from the decomposition (7.120) of the incident field $\mathbf{E}_{\text{inc}}(\mathbf{r})$, which is finite at $\mathbf{r} = \mathbf{0}$.

The problem now is to determine the four sequences of coefficients $A_{\mathrm{e}mn}$, $A_{\mathrm{o}mn}$, $B_{\mathrm{e}mn}$, and $B_{\mathrm{o}mn}$ appearing in the expansion (7.120). To do this, we appeal to the orthogonality of the vector spherical harmonics $\mathbf{M}_{\mathrm{e}mn}$, $\mathbf{M}_{\mathrm{o}mn}$, $\mathbf{N}_{\mathrm{e}mn}$, and $\mathbf{N}_{\mathrm{o}mn}$, i.e., whatever two harmonics $\mathbf{X}$ and $\mathbf{Y}$ are chosen from the set

$$\Big\{\mathbf{M}_{\mathrm{e}mn},\,\mathbf{M}_{\mathrm{o}mn},\,\mathbf{N}_{\mathrm{e}mn},\,\mathbf{N}_{\mathrm{o}mn},\ m,n\in\mathbb{Z}^+,\ n\geq m\Big\}\ ,$$

we have

$$\mathbf{X}\neq\mathbf{Y} \quad\Longrightarrow\quad \int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\mathbf{X}(r,\theta,\varphi)\cdot\mathbf{Y}(r,\theta,\varphi)\sin\theta\,\mathrm{d}\theta\mathrm{d}\varphi = 0\ . \tag{7.121}$$

Expressions for the unknown coefficients follow from this remarkable property. For example, $A_{\mathrm{e}mn}$ can be calculated from the relation

$$A_{\mathrm{e}mn} = \frac{\displaystyle\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\mathbf{E}_{\text{inc}}(r,\theta,\varphi)\cdot\mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi)\sin\theta\,\mathrm{d}\theta\mathrm{d}\varphi}{\displaystyle\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi)\cdot\mathbf{M}_{\mathrm{e}mn}(r,\theta,\varphi)\sin\theta\,\mathrm{d}\theta\mathrm{d}\varphi}\ , \tag{7.122}$$

with similar results for $A_{\mathrm{o}mn}$, $B_{\mathrm{e}mn}$, and $B_{\mathrm{o}mn}$. Pursuing the mathematics in this direction, it can be shown that all the coefficients $A_{\mathrm{e}mn}$ and $B_{\mathrm{o}mn}$ are zero, and that the remaining coefficients $A_{\mathrm{o}mn}$ and $B_{\mathrm{e}mn}$ are non-zero only when $m = 1$, in which case

$$A_{\mathrm{o}1n} = \mathrm{i}^n E_0\frac{2n+1}{n(n+1)}\ ,\quad B_{\mathrm{e}1n} = -\mathrm{i}^{n+1}E_0\frac{2n+1}{n(n+1)} = -\mathrm{i}A_{\mathrm{o}1n}\ . \tag{7.123}$$

Then introducing the notation

$$E_n = A_{\mathrm{o}1n} = \mathrm{i}^n E_0 \frac{2n+1}{n(n+1)} ,$$

the incident electric field $\mathbf{E}_{\mathrm{inc}}(r,\theta,\varphi)$ has the following expansion relative to the basis of vector spherical harmonics:

$$\boxed{\mathbf{E}_{\mathrm{inc}}(r,\theta,\varphi) = \sum_{n=1}^{\infty} E_n \left[\mathbf{M}^{j2}_{\mathrm{o}1n}(r,\theta,\varphi) - \mathrm{i}\mathbf{N}^{j2}_{\mathrm{e}1n}(r,\theta,\varphi) \right]} \tag{7.124}$$

in which the superscript $j2$ on $\mathbf{M}^{j2}_{\mathrm{o}1n}$ and $\mathbf{N}^{j2}_{\mathrm{e}1n}$ indicates that the functions $z_n(kr)$ appearing in the expressions (7.115) and (7.116) for these two harmonics must be taken equal to the spherical Bessel functions $j_n(k_2 r)$ expressed in medium 2.

Fourth Stage: Expanding the Internal and Scattered Fields in Terms of Vector Spherical Harmonics

We now have the expansion of the incident electromagnetic field $\left\{\mathbf{E}_{\mathrm{inc}}(\mathbf{r}), \mathbf{H}_{\mathrm{inc}}(\mathbf{r})\right\}$ in terms of the vector spherical harmonics:

$$\begin{cases} \mathbf{E}_{\mathrm{inc}}(r,\theta,\varphi) = \displaystyle\sum_{n=1}^{\infty} E_n \left[\mathbf{M}^{j2}_{\mathrm{o}1n}(r,\theta,\varphi) - \mathrm{i}\mathbf{N}^{j2}_{\mathrm{e}1n}(r,\theta,\varphi) \right] , \\ \mathbf{H}_{\mathrm{inc}}(r,\theta,\varphi) = \dfrac{\nabla \times \mathbf{E}_{\mathrm{inc}}(r,\theta,\varphi)}{\mathrm{i}\omega\mu_2} = \dfrac{k_2}{\mathrm{i}\omega\mu_2} \displaystyle\sum_{n=1}^{\infty} E_n \left[\mathbf{N}^{j2}_{\mathrm{o}1n}(r,\theta,\varphi) - \mathrm{i}\mathbf{M}^{j2}_{\mathrm{e}1n}(r,\theta,\varphi) \right] . \end{cases} \tag{7.125}$$

Since Maxwell's equations are linear in the fields, we may deduce that the expansions in terms of vector spherical harmonics of the field $\left\{\mathbf{E}_1(r,\theta,\varphi), \mathbf{H}_1(r,\theta,\varphi)\right\}$ within the spherical particle and the field $\left\{\mathbf{E}_{\mathrm{sca}}(r,\theta,\varphi), \mathbf{H}_{\mathrm{sca}}(r,\theta,\varphi)\right\}$ scattered by it will be made up of exactly the same modes as the incident electromagnetic field $\left\{\mathbf{E}_{\mathrm{inc}}(r,\theta,\varphi), \mathbf{H}_{\mathrm{inc}}(r,\theta,\varphi)\right\}$ (thinking of a vector spherical harmonic as the analogue of a mode in standard frequency analysis). It is thus perfectly justified to expand the internal and scattered fields $\left\{\mathbf{E}_1(r,\theta,\varphi), \mathbf{H}_1(r,\theta,\varphi)\right\}$ and $\left\{\mathbf{E}_{\mathrm{sca}}(r,\theta,\varphi), \mathbf{H}_{\mathrm{sca}}(r,\theta,\varphi)\right\}$ in the following way:

$$\boxed{\begin{cases} \mathbf{E}_1(r,\theta,\varphi) = \displaystyle\sum_{n=1}^{\infty} E_n \left[c_n \mathbf{M}^{j1}_{\mathrm{o}1n}(r,\theta,\varphi) - \mathrm{i}d_n \mathbf{N}^{j1}_{\mathrm{e}1n}(r,\theta,\varphi) \right] , \\ \mathbf{H}_1(r,\theta,\varphi) = \dfrac{k_1}{\mathrm{i}\omega\mu_1} \displaystyle\sum_{n=1}^{\infty} E_n \left[c_n \mathbf{N}^{j1}_{\mathrm{o}1n}(r,\theta,\varphi) - \mathrm{i}d_n \mathbf{M}^{j1}_{\mathrm{e}1n}(r,\theta,\varphi) \right] , \end{cases}} \tag{7.126}$$

$$\boxed{\begin{cases} \mathbf{E}_{\text{sca}}(r,\theta,\varphi) = \displaystyle\sum_{n=1}^{\infty} E_n \Big[-b_n \mathbf{M}^{h2}_{\text{o}1n}(r,\theta,\varphi) + \text{i}a_n \mathbf{N}^{h2}_{\text{e}1n}(r,\theta,\varphi) \Big] , \\ \mathbf{H}_{\text{sca}}(r,\theta,\varphi) = \dfrac{k_2}{\text{i}\omega\mu_2} \displaystyle\sum_{n=1}^{\infty} E_n \Big[-b_n \mathbf{N}^{h2}_{\text{o}1n}(r,\theta,\varphi) + \text{i}a_n \mathbf{M}^{h2}_{\text{e}1n}(r,\theta,\varphi) \Big] , \end{cases}} \tag{7.127}$$

and it remains to determine the four sequences of complex coefficients a_n, b_n, c_n, and d_n. Note in passing that the functions $z_n(kr)$ appearing in the internal field must be taken as the functions $j_n(k_1 r)$, because $y_n(\rho) \to -\infty$ when $\rho \to 0$. Furthermore, since the scattered field has the asymptotic form $\exp(\text{i}k_2 r)/(-\text{i}k_2 r)$ in the radiative region [see (7.15) in Sect. 7.2.4], it can be shown that its functions $z_n(kr)$ must be the spherical Hankel functions $h_n(k_2 r)$ defined by $h_n(k_2 r) = j_n(k_2 r) + \text{i}y_n(k_2 r)$.

The values of the four sequences of complex coefficients a_n, b_n, c_n, and d_n follow from continuity of the tangential components, i.e., in the directions $\mathbf{e}_\theta$ and $\mathbf{e}_\varphi$, of the electric and magnetic fields at the interface between the particle and the host medium. If a is the particle radius, these continuity conditions can be written in the form

$$\forall\theta\,,\ \forall\varphi\,, \quad \begin{cases} \mathbf{E}_1(r=a,\theta,\varphi)\cdot\mathbf{e}_\theta = \Big[\mathbf{E}_{\text{inc}}(r=a,\theta,\varphi) + \mathbf{E}_{\text{sca}}(r=a,\theta,\varphi)\Big]\cdot\mathbf{e}_\theta\ , \\ \mathbf{E}_1(r=a,\theta,\varphi)\cdot\mathbf{e}_\varphi = \Big[\mathbf{E}_{\text{inc}}(r=a,\theta,\varphi) + \mathbf{E}_{\text{sca}}(r=a,\theta,\varphi)\Big]\cdot\mathbf{e}_\varphi\ , \\ \mathbf{H}_1(r=a,\theta,\varphi)\cdot\mathbf{e}_\theta = \Big[\mathbf{H}_{\text{inc}}(r=a,\theta,\varphi) + \mathbf{H}_{\text{sca}}(r=a,\theta,\varphi)\Big]\cdot\mathbf{e}_\theta\ , \\ \mathbf{H}_1(r=a,\theta,\varphi)\cdot\mathbf{e}_\varphi = \Big[\mathbf{H}_{\text{inc}}(r=a,\theta,\varphi) + \mathbf{H}_{\text{sca}}(r=a,\theta,\varphi)\Big]\cdot\mathbf{e}_\varphi\ , \end{cases} \tag{7.128}$$

and expanding in terms of the different modes $n \geq 1$ present in the expressions (7.125)–(7.127) for the three electromagnetic fields, these imply the following system:

$$\begin{cases} h_n(x) b_n + j_n(mx) c_n = j_n(x)\ , \\ m\big[\rho h_n(\rho)\big]'(x) a_n + \big[\rho\, j_n(\rho)\big]'(mx) d_n = m\big[\rho\, j_n(\rho)\big]'(x)\ , \\ \mu_1 h_n(x) a_n + \mu_2 m j_n(mx) d_n = \mu_1 j_n(x)\ , \\ \mu_1 \big[\rho h_n(\rho)\big]'(x) b_n + \mu_2 \big[\rho\, j_n(\rho)\big]'(mx) c_n = \mu_1 \big[\rho\, j_n(\rho)\big]'(x)\ . \end{cases} \tag{7.129}$$

This is a linear system from which analytic expressions for the coefficients a_n, b_n, c_n, and d_n can be extracted. In (7.129), we have introduced the following notation:

- The quantity x defined by $x = k_2 a$, a dimensionless parameter commonly called the size parameter relative to the host medium (labelled 2). x is real, because the host medium is assumed non-absorbent.
- The quantity m defined by $m = m_1/m_2 = k_1/k_2$, also dimensionless. The product mx is thus equal to $k_1 a$, the size parameter relative to the particle (labelled 1). This quantity may be complex, if the particle is made from a material absorbing the radiation.

The Mie problem, that is, the analytic determination of the electromagnetic field $\{\mathbf{E}_1(r,\theta,\varphi),\mathbf{H}_1(r,\theta,\varphi)\}$ within the spherical particle and the electromagnetic field $\{\mathbf{E}_{\rm sca}(r,\theta,\varphi),\mathbf{H}_{\rm sca}(r,\theta,\varphi)\}$ scattered by it, has thus been formally solved. Having said that, the expressions for the coefficients a_n, b_n, c_n, and d_n are generally rather complicated. A case that is often encountered in the literature concerns situations where the particle and the host medium have the same magnetic permeability. The coefficients a_n and b_n of the scattered field $\{\mathbf{E}_{\rm sca}(r,\theta,\varphi),\mathbf{H}_{\rm sca}(r,\theta,\varphi)\}$ then simplify somewhat and can be written in the following form:

$$\boxed{\begin{aligned} a_n &= \frac{m\psi_n(mx)\psi_n'(x)-\psi_n(x)\psi_n'(mx)}{m\psi_n(mx)\xi_n'(x)-\xi_n(x)\psi_n'(mx)}\,, \\ b_n &= \frac{\psi_n(mx)\psi_n'(x)-m\psi_n(x)\psi_n'(mx)}{\psi_n(mx)\xi_n'(x)-m\xi_n(x)\psi_n'(mx)}\,, \end{aligned}} \tag{7.130}$$

where the functions $\psi_n(\rho)$ and $\xi_n(\rho)$ are the Ricatti–Bessel functions defined by

$$\psi_n(\rho)=\rho\, j_n(\rho)\,, \qquad \xi_n(\rho)=\rho\, h_n(\rho)\,. \tag{7.131}$$

From (7.130), it follows that the two sequences of scattering coefficients a_n and b_n are functions only of the two dimensionless parameters introduced above, viz., the size parameter x (real) and the complex optical index contrast m (possibly complex).

References

1. R. Carminati: Introduction to radiative transfer. In: *Microscale and Nanoscale Heat Transfer*, ed. by S. Volz, Topics in Applied Physics Vol. 107, Springer, Berlin Heidelberg, 2007
2. J. Taine, J.-P. Petit: *Transferts thermiques. Mécanique des fluides anisothermes. Cours et données de base*, Dunod, Paris, 1995
3. C.F. Bohren, D.R. Huffman: *Absorption and Scattering of Light by Small Particles*, Wiley-VCH, Weinheim, 1983
4. J.D. Jackson: *Classical Electrodynamics*, John Wiley, New York, 1975
5. M. Born, E. Wolf: *Principles of Optics. Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 6th edn., Pergamon Press, Oxford, 1980
6. M.F. Modest: *Radiative Heat Transfer*, McGraw-Hill, New York, 1993
7. G.A. Mie: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen, Annalen der Physik **25**, 377–445 (1908)
8. H.C. Van de Hulst: *Light Scattering by Small Particles*, Dover Publications, New York, 1981
9. W.J. Wiscombe: Mie scattering calculations: Advances in technique and fast, vector speed computer codes, technical note, National Center for Atmospheric Research, Boulder, CO, USA (1996)
10. H. Du: Mie scattering calculation, Applied Optics **43**, 1951–1956 (2004)
11. C.L. Tien, B.L. Drolen: Thermal radiation in particulate media with dependent and independent scattering. In: *Annual Review of Numerical Fluid Mechanics and Heat Transfer*, Vol. 1, Hemisphere, New York, 1987
12. H.S. Chu, A.J. Stretton, C.L. Tien: Radiative heat transfer in ultra-fine powder insulations, International Journal of Heat and Mass Transfer **31**, 1627–1634 (1988)

13. M. Kaviany: *Principles of Heat Transfer in Porous Media*, 2nd edn., Springer, New York, 1995
14. D. Baillis, J.-F. Sacadura: Thermal radiation properties of dispersed media: Theoretical prediction and experimental characterization, Journal of Quantitative Spectroscopy and Radiative Transfer **67**, 327–363 (2000)
15. E.D. Palik (Ed.): *Handbook of Optical Constants of Solids*, Vol. 2, Academic Press, San Diego, 1991
16. E.M. Purcell, C.R. Pennypacker: Scattering and absorption by nonspherical dielectric grains, Astrophysical Journal **186**, 705–714 (1973)
17. T. Kozasa, J. Blum, T. Mukai: Optical properties of dust aggregates I. Wavelength dependence, Astronomy and Astrophysics **263**, 423–432 (1992)
18. H. Okamoto, Y.-L. Xu: Light scattering by irregular interplanetary dust particles, Earth Planets Space **50**, 577–585 (1998)
19. B.T. Draine: The discrete dipole approximation and its application to interstellar graphite grains, Astrophysical Journal **333**, 848–872 (1988)
20. B.T. Draine, J. Goodman: Beyond Clausius–Mossotti: Wave propagation on a polarizable point lattice and the discrete dipole approximation, Astrophysical Journal **405**, 685–697 (1993)
21. B.T. Draine, P.J. Flatau: Discrete dipole approximation for scattering calculations, Journal of the Optical Society of America A **11**, 1491–1499 (1994)
22. F. Enguehard, S. Lallich: L'Approximation dipolaire discrète: Théorie et premiers résultats, technical note, Commissariat à l'Energie Atomique, Centre du Ripault, Monts, France (2008) (available on request)
23. D. Lemonnier: Solution of the Boltzmann equation for phonon transport. In: *Microscale and Nanoscale Heat Transfer*, ed. by S. Volz, Topics in Applied Physics Vol. 107, Springer, Berlin Heidelberg, 2007
24. W.T. Doyle: Optical properties of a suspension of metal spheres, Physical Review B **39**, 9852–9858 (1989)
25. H. Okamoto: Light scattering by clusters: The a1 term method, Optical Review **2**, 407–412 (1995)
26. C.E. Dungey, C.F. Bohren: Light scattering by nonspherical particles: A refinement to the coupled dipole method, Journal of the Optical Society of America A **8**, 81–87 (1991)
27. R. Koch, R. Becker: Evaluation of quadrature schemes for the discrete ordinate method, Journal of Quantitative Spectroscopy and Radiative Transfer **84**, 423–435 (2004)
28. D. Rochais, G. Domingues, F. Enguehard: Transferts thermiques dans les isolants microporeux, Chocs (scientific and technical review of the Direction des Applications Militaires of the CEA) **33**, 37–43 (2006)
29. S. Lallich, F. Enguehard, D. Baillis: Propriétés optiques et radiatives de matrices nanoporeuses de silice, Congrès Français de Thermique SFT-2007 (Ile des Embiez, France, 29 May to 1 June 2007), 483–488 (2007)
30. A. Legrand: *The Surface Properties of Silicas*, Wiley, New York, 1998

Chapter 8
Thermal Conductivity of Nanofluids

Pawel Keblinski

Nanofluids (colloidal suspensions of solid nanoparticles) sparked excitement as well as controversy. In particular, a number of researches reported dramatic increases of thermal conductivity with small nanoparticle loading, while others showed moderate increases consistent with the effective medium theories on well-dispersed conductive spheres. Here we discuss potential mechanisms that were put forward in order to understand nanoflouid thermal conductivity and demonstrate that particle aggregation is currently the only physically reasonable mechanism that can explain the majority of the experimental data.

8.1 Introduction

Cooling is one of the most important technical challenges facing numerous diverse industries, including microelectronics, transportation, solid-state lighting, and manufacturing. Developments driving the increased thermal loads that require advances in cooling include faster speeds (in the multi-GHz range) and smaller features (to < 100 nm) for microelectronic devices, higher power engines, and brighter optical devices. The conventional method for increasing heat dissipation is to increase the area available for exchanging heat with a heat transfer fluid, but this approach requires an undesirable increase in the size of thermal management system. There is therefore an urgent need for new and innovative coolants with improved performance. About a decade ago a novel concept of nanofluids, i.e., heat transfer fluids containing suspensions of nanoparticles, was proposed as a means of meeting these challenges [1].

Nanofluids are solid–liquid composite materials consisting of solid nanoparticles or nanofibers, with sizes typically on the order of 1–100 nm, suspended in a liquid. Nanofluids have attracted great interest recently due to reports of greatly enhanced thermal properties at low volume fractions. For example, a small amount (less than 1% volume fraction) of copper nanoparticles and carbon nanotubes dispersed

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 213
DOI 10.1007/978-3-642-04258-4_8, © Springer-Verlag Berlin Heidelberg 2009

in ethylene glycol and oil, respectively, was reported to increase the inherently poor thermal conductivity of the liquid by 40% and 150%, respectively [2–5]. Conventional suspensions of well-dispersed particles require high concentrations ($> 10\%$) of particles to achieve such enhancement; problems of rheology and stability, which are amplified at high concentrations, preclude the widespread use of conventional slurries as heat transfer fluids. In several cases, the observed enhancement in thermal conductivity of nanofluids is orders-of-magnitude larger than predicted by well-established theories of dispersed particles [6–12].

8.2 Excitement, Controversy, and New Physics

The large thermal conductivity enhancements reported by experiment led to excitement but also to controversy. The origin of the excitement was that the measured thermal conductivity was often much larger than that predicted by well-established effective medium theories under the assumption of well-dispersed particles. For a system of well dispersed particles, in the limit of low particle volume fraction ϕ, and with much higher particle conductivity $\Lambda_{\rm NP}$ than fluid conductivity $\Lambda_{\rm fluid}$, the effective medium (Maxwell) theory for vanishing interfacial thermal resistance leads to the relation [13, 14]

$$\frac{\Lambda_{\rm NF}}{\Lambda_{\rm fluid}} = 1 + 3\phi \, . \tag{8.1}$$

Equation (8.1) predicts only moderate thermal transport increase, independently of the conductivity of the filler. As already discussed, many experiments reported much larger increases.

This mismatch between predicted and observed values led not only to excitement but also to controversy, since in some experiments, the measured thermal conductivity was not very large and consistent with the prediction of (8.1) [15–19]. This initially led to the belief that some experiments had to be wrong, particularly considering that different research groups often obtained different results for what were presumably the same nanofluids.

Further excitement and controversy were associated with potential mechanisms put forward to explain the 'anomalous' experimental observations. This discussion was initiated by Keblinski et al. [20]. In the following we will review the key mechanism analyzed and discussed in the literature.

8.2.1 Brownian Motion

Motivated by the 'unusual' thermal conductivity enhancements, a number of researchers promoted the concept of Brownian motion induced micro- or nanoconvection [21–23]. They argued in different ways that each Brownian particle generates a long-range velocity field in the surrounding fluid, akin to that present around a par-

ticle moving with a constant velocity, which decays approximately as the inverse of the distance from the particle center. The ability of large volumes of fluid dragged by the nanoparticles to carry substantial amounts of heat was credited for the large thermal conductivity increases in nanofluids.

A key weakness of this argument is that the thermal diffusivity D_T of the base fluid, which measures the rate of the heat flow via thermal conduction, is several orders of magnitude larger than nanoparticle diffusivity D_{NP}, which measures the rate of mass motion due to nanoparticle diffusion, whence the magnitude of possible nanoconvection effects is negligible [24]. Furthermore, the velocity field around a Brownian particle is much shorter-ranged than that around a particle moving with a constant velocity [25]. The low estimates for the contribution to thermal transport of Brownian motion were supported by the results of molecular dynamics simulations [24,26], and by the results of experimental measurements on well-dispersed spherical nanoparticle fluids [16–19,27], all showing thermal conductivity enhancements (positive and negative) that agree with (8.1). In a direct experimental investigation, the density effects associated with the postulated nanoconvection [23] were experimentally tested with lighter silica and Teflon particles, and were shown to be incorrect [28]. The nanoconvection velocities were further shown to be of the order of thermophoretic velocities, which for most nanofluids were insignificant (as low as 10^{-9} m/s). Even for sub-nanometer clusters, as evidenced from molecular dynamics simulations, the thermophoretic velocities are exceedingly small, and this effectively precludes a discernible contribution to the nanofluid thermal conductivity from any conceivable nanoconvection mechanism [29].

8.2.2 Interfacial Liquid Layer

Considering that the molecular structure of liquid at the solid interface is more ordered, possibilities of larger thermal conductivity of this ordered liquid layer and 'tunneling' of heat carrying phonons from one particle to another were put forward [20]. The subsequent molecular dynamics simulation work concluded that those mechanisms do not contribute significantly to heat transfer [30]. For strong solid–liquid interactions, typical of those in nanofluids with metallic nanoparticles, a percolating network of amorphous-like fluid structures can emerge which can facilitate additional thermal conduction paths [31, 32]. However, a discernible increase in thermal conductivity is possible only for exceedingly small colloidal particles (limited to few tens of atoms). Experiments have shown that the structured interfacial fluid layers are limited to a few molecular spacings from a solid surface [33], which makes them inadequate to influence the thermal transport in common nanofluids. Furthermore, interpretation of the cooling rates of Au nanoparticles suspended in water and organic solvents does not appear to require any unusual thermophysical properties of the surrounding liquid to explain the experimental results [34].

8.2.3 Interfacial Thermal Resistance

As described above, the liquid layering does not generate thermal conductivity increases. There is, however, a well-known interfacial effect degrading thermal conductivity: the thermal interfacial resistance R_k, most commonly defined via its inverse, the interfacial conductance $G = 1/R_k$. G is related to the heat flux J_Q and the temperature drop ΔT at the interface via

$$J_Q = G\Delta T \ . \tag{8.2}$$

A simple measure of the relative importance of the interfacial resistance in the overall heat flow in a composite can be obtained from the equivalent thickness h, defined as the distance over which the temperature drop is the same as at the interface. This thickness is given by the ratio of the fluid thermal conductivity Λ_{fluid} to the interfacial conductance, i.e., $h = \Lambda_{\text{fluid}}/G$.

The interfacial thermal resistance effect on well-dispersed spherical particle composites can be estimated by effective medium theory to be [15]

$$\frac{\Lambda_{\text{NF}}}{\Lambda_{\text{fluid}}} - 1 = 3\phi\frac{\gamma - 1}{\gamma + 2} \ , \tag{8.3}$$

where γ is the ratio of the particle radius to the equivalent matrix thickness h. According to (8.3), when the particle radius becomes equal to the equivalent matrix thickness ($\gamma = 1$), there is no enhancement at all, while for larger interfacial resistance ($\gamma < 1$), the addition of particles decreases the thermal conductivity of the composite. This effect can explain the decrease of thermal conductivity of nanofluids below the base fluid value [18, 19] and much lower than expected thermal conductivity increases of carbon nanotube nanofluids and carbon nanotube polymer composites [35, 36].

8.2.4 Near Field Radiation

Somewhat less popular was the idea of anomalously high radiative energy transfers between nanoparticles [37]. Since the classical radiation theory does not predict radiative transfer of significance for nanofluids, molecular dynamics simulations of thermal energy exchange between two silica nanoparticles were carried out, resulting in the prediction of enormous thermal conductance due to near-field interactions when particles were closely separated [37]. However, the thermal conductance obtained from molecular dynamics was significantly larger than that obtained under the assumption that all the thermal energy is exchanged between two nanoparticles within a single atomic vibration period. Further analysis demonstrated that near-field interactions do not affect the nanofluid thermal conductivity [38].

8.2.5 Particle Clustering

Maxwell's expression, the limiting case of which is given by (8.1), corresponds to the situation of well-dispersed nanoparticles. Such structure is the least efficient from the point of view of thermal transport enhancement since conductive particles are separated from each other by low conductivity liquid. Aggregation of particles into sparse clusters, or ideally into linear chains, leads to extended and highly conductive paths for the heat flow. In fact, there is a well established understanding of the impact of morphology on thermal conductivity of a composite. Depending on the morphology, the conductivity can vary greatly, even at the same volume fraction of the components. It is generally accepted that the conductivity has to fall between so-called Hashin–Shtrikman (HS) bounds, obtained under an effective medium analysis [39] without any restriction on the volume fraction. The HS bounds are given by

$$\Lambda_{\text{fluid}}\left[1+\frac{3\phi(\Lambda_{\text{NP}}-\Lambda_{\text{fluid}})}{3\Lambda_{\text{fluid}}+(1-\phi)(\Lambda_{\text{NP}}-\Lambda_{\text{fluid}})}\right]\leq\Lambda_{\text{NF}}\leq\left[1-\frac{3(1-\phi)(\Lambda_{\text{NP}}-\Lambda_{\text{fluid}})}{3\Lambda_{\text{NP}}-\phi(\Lambda_{\text{NP}}-\Lambda_{\text{fluid}})}\right]\Lambda_{\text{NP}}\,. \tag{8.4}$$

In the inequality given by (8.4), the lower (Maxwell) bound corresponds physically to a set of well-dispersed nanoparticles in a fluid matrix, while the upper limit corresponds to large pockets of fluid separated by linked, chain-forming or clustered nanoparticles. From the point of view of circuit analysis, the lower HS limit lies closer to conductors connected in series, while the upper limit lies closer to conductors in a parallel mode. The HS bounds do not give a precise mechanism of thermal conductance, but set the most restrictive limits based on knowledge of the volume fraction alone. It is relatively well-known that a large number of experimental data on solid composites, and to a lesser extent the data on liquid mixtures, fall between the HS bounds [29]. An unbiased estimate (favoring neither series nor parallel modes) predicts thermal conductivity values that lie between the upper and lower HS bounds [40].

The evidence from scanning electron microscopy (SEM) points to the existence of partial clustering and chain-like linear aggregation [10–12, 41]. Viscosity data on nanofluids has also shown an anomalous increase as compared to the Einstein model for well dispersed particles [42]. A large increase in the viscosity is another indication of aggregation in the nanofluids. Interestingly, for a carbon nanotube suspension, an effective medium theory accounting for the very high aspect ratio of the nanotubes [14], predicts thermal conductivity enhancements that are in fact well above the reported values. This was attributed to a significant interfacial resistance to heat flow between the carbon nanotubes and the fluid [35].

8.3 Discussion

Clear experimental evidence for extensive clustering observed in nanofluids indicates that the expected thermal conductivity can fall within a wide range bounded by (8.4) and strongly depends on the actual aggregation. Recently, using a multilevel effective medium theory, it was demonstrated that the thermal conductivity of nanofluids can be significantly enhanced by the aggregation of nanoparticles into chain-like clusters, and this enhancement can be quite dramatic for large, but sparse clusters [43]. Predictions of this aggregation model were in excellent agreement with detailed numerical calculations on model nanofluids involving fractal clusters [43]. Thus, allowing for clustering effects dramatically broadens the thermal conductivity range that is consistent with the effective medium theory [29]. Quite strikingly, it was demonstrated that the vast majority of experimental results fall within these bounds, supporting the classical nature of thermal conduction behavior in nanofluids [44].

The increase in the relative thermal conductivity of nanofluids with temperature [4] is another example of anomalous behavior that cannot be explained on the assumption of well-dispersed particles. Similarly, the increase in the thermal conductivity with decreasing particle size cannot be explained if the particles are well dispersed. The probability of aggregation increases with increasing temperature and decreasing particle size [45]. Therefore the apparent contradictions between experiment and theory, such as particle size effects, can be resolved by weighing in the ability of nanoparticles to form linear clusters. Furthermore, the temperature dependence is not as strong as it was previously believed to be, with recent experiments showing a similar variation for both nanofluids and the base fluid [19, 29]. This implies that the mechanism for increase in the thermal conductivity of water (presumably from the hydrogen bonded structures) is partly responsible for the thermal conductivity increase in nanofluids as well. Conversely, it is reasonable to expect a decrease in the nanofluid thermal conductivity for a base fluid that has a negative change in conductivity with increasing temperature.

It remains a challenge to accurately identify and manipulate the cluster configuration to modify the thermal transport properties of a nanofluid. The two characterization techniques, DLS and SEM, have limitations in assessing the structure of nanoparticles. DLS measurements are limited to dilute suspensions ($\phi < 1\%$) for most nanofluids, while SEM imaging can be performed only after drying the base fluid. While the science of making well-dispersed colloids has reached a fair level of maturity, attempts to generate targeted nanoparticle configurations are still in an evolving phase.

The fact that significant aggregation is required to obtain substantial increases in thermal transport has an important consequence for the potential application of such fluids in flow-based cooling, which is the most important benefit from the technological point of view. It is well known that aggregation into sparse but large clusters increases fluid viscosity. Such increases become very dramatic when the aggregates start to touch each other, which can occur at very low volume fractions, as low as 0.2% [46]. Therefore, the same aggregate structures that are most beneficial to the

thermal transport are also the most detrimental to the fluid flow characteristics. Future research should therefore address the issue of optimizing nanofluid structure with the best combination of thermal conductivity and viscosity.

References

1. Choi, S.U.S.: Enhancing thermal conductivity of fluids with nanoparticles. In: Siginer, D.A., Wang, H.P., and Div., F.E. (Eds.) *Developments and Applications of Non-Newtonian Flows*, American Society of Mechanical Engineers, New York (1995) pp. 99–105
2. Choi, S.U.S., Zhang, Z.G., Yu, W., Lockwood, F.E., and Grulke, E.A.: Anomalous thermal conductivity enhancement in nanotube suspensions. Appl. Phys. Lett. **79**, 2252–2254 (2001)
3. Eastman, J.A., Choi, S.U.S., Li, S., Yu, W., and Thompson, L.J.: Anomalously increased effective thermal conductivities of ethylene glycol-based nanofluids containing copper nanoparticles. Appl. Phys. Lett. **78**, 718–720 (2001)
4. Das, S.K., Putra, N., Thiesen, P., and Roetzel, W.: Temperature dependence of thermal conductivity enhancement for nanofluids. J. Heat Transfer **125**, 567–574 (2003)
5. Patel, H.E., Das, S.K., Sundararajan, T., Nair, A.S., George, B., and Pradeep, T.: Thermal conductivities of naked and monolayer protected metal nanoparticle based nanofluids: Manifestation of anomalous enhancement and chemical effects. Appl. Phys. Lett. **83**, 2931–2933 (2003)
6. Chopkar, M., Das, P.K., and Manna, I.: Synthesis and characterization of nanofluid for advanced heat transfer applications. Scripta Materialia **55**, 549–552 (2006)
7. Chopkar, M., Kumar, S., Bhandari, D.R., Das, P.K., and Manna, I.: Development and characterization of Al_2Cu and Ag_2Al nanoparticle dispersed water and ethylene glycol based nanofluid. Materials Science and Engineering: B **139**, 141–148 (2007)
8. Hong, K.S., Hong, T.-K., and Yang, H.-S.: Thermal conductivity of Fe nanofluids depending on the cluster size of nanoparticles. Appl. Phys. Lett. **88**, 031901 (2006)
9. Kang, H.U., Kim, S.H., and Oh, J.M.: Estimation of thermal conductivity of nanofluid using experimental effective particle volume. Experimental Heat Transfer **19**, 181–191 (2006)
10. Murshed, S.M.S., Leong, K.C., and Yang, C.: Enhanced thermal conductivity of TiO_2–water based nanofluids Int. J. Therm. Sci. **44**, 367–373 (2005)
11. Zhu, H., Zhang, C., Liu, S., Tang, Y., and Yin, Y.: Effects of nanoparticle clustering and alignment on thermal conductivities of Fe_3O_4 aqueous nanofluids. Appl. Phys. Lett. **89**, 023123 (2006)
12. Zhu, H.T., Zhang, C.Y., Tang, Y.M., and Wang, J.X.: Novel synthesis and thermal conductivity of CuO nanofluid. J. Phys. Chem. C **111**, 1646–1650 (2007)
13. Maxwell, J.C.: *A Treatise on Electricity and Magnetism*, II edn. Clarendon, Oxford (1881)
14. Nan, C.-W., Birringer, R., Clarke, D.R., and Gleiter, H.: Effective thermal conductivity of particulate composites with interfacial thermal resistance. J. Appl. Phys. **81**, 6692–6699 (1997)
15. Keblinski, P., Eastman, J.A., and Cahill, D.G.: Nanofluids for thermal transport. Materials Today **8**, 36–44 (2005)
16. Putnam, S.A., Cahill, D.G., Braun, P.V., Ge, Z., and Shimmin, R.G.: Thermal conductivity of nanoparticle suspensions. J. Appl. Phys. **99**, 084308 (2006)
17. Venerus, D.C., Kabadi, M.S., Lee, S., and Perez-Luna, V.: Study of thermal transport in nanoparticle suspensions using forced Rayleigh scattering. J. Appl. Phys. **100**, 094310 (2006)
18. Zhang, X., Gu, H., and Fujii, M.: Effective thermal conductivity and thermal diffusivity of nanofluids containing spherical and cylindrical nanoparticles. J. Appl. Phys. **100**, 044325 (2006)
19. Zhang, X., Gu, H., and Fujii, M.: Experimental study on the effective thermal conductivity and thermal diffusivity of nanofluids. Int. J. Thermophysics **27**, 569–580 (2006)

20. Keblinski, P., Phillpot, S.R., Choi, S.U.S., and Eastman, J.A.: Mechanisms of heat flow in suspensions of nano-sized particles (nanofluids). Int. J. Heat and Mass Transfer **45**, 855–863 (2002)
21. Jang, S.P., and Choi, S.U.S.: Role of Brownian motion in the enhanced thermal conductivity of nanofluids. Appl. Phys. Lett. **84**, 4316–4318 (2004)
22. Koo, J., and Kleinstreuer, C.: A new thermal conductivity model for nanofluids. J. Nanoparticle Res. **6**, 577–588 (2004)
23. Prasher, R., Bhattacharya, P., and Phelan, P.E.: Thermal conductivity of nanoscale colloidal solutions (nanofluids). Phys. Rev. Lett. **94**, 025901 (2005)
24. Evans, W., Fish, J., and Keblinski, P.: Role of Brownian motion hydrodynamics on nanofluid thermal conductivity. Appl. Phys. Lett. **88**, 093116 (2006)
25. Keblinski, P., and Thomin: Hydrodynamic field around a Brownian particle, Phys. Rev. E **73**, Rapid Communication 010502 (2006)
26. Vladkov, M., and Barrat, J.-L.: Modeling transient absorption and thermal conductivity in a simple nanofluid. Nano Lett. **6**, 1224–1228 (2006)
27. Rusconi, R., Rodari, E., and Piazza, R.: Optical measurements of the thermal properties of nanofluids. Appl. Phys. Lett **89**, 261916 (2006)
28. Eapen, J., Williams, W.C., Buongiorno, J., Hu, L.-W., Yip, S., Rusconi, R., and Piazza, R.: Mean-field versus microconvection effects in nanofluid thermal conduction. Phys. Rev. Lett. **99**, 095901 (2007)
29. Eapen, J., Buongiorno, J., Hu, L.-W., Yip, S., Rusconi, R., and Piazza, R.: Mean-field bounds and the classical nature of thermal conduction in nanofluids. Manuscript under preparation (2007)
30. Xue, L., Keblinski, P., Phillpot, S.R., Choi, S.U.S., and Eastman, J.A.: Effect of liquid layering at the liquid–solid interface on thermal transport. Int. J. Heat and Mass Transfer **47**, 4277–4283 (2004)
31. Eapen, J., Li, J., and Yip, S.: Beyond the Maxwell limit: Thermal conduction in nanofluids with percolating fluid structures. In press, Phys. Rev. E. `arXiv:0707.2164v1` (2007)
32. Eapen, J., Li, J., and Yip, S.: Mechanism of thermal transport in dilute nanocolloids. Phys. Rev. Lett. **98**, 028302 (2007)
33. Yu, C.-J., Richter, A.G., Kmetko, J., Dugan, S.W., Datta, A., and Dutta, P.: Structure of interfacial liquids: X-ray scattering studies. Phys. Rev. E **63**, 021205 (2001)
34. Wilson, O.M., Hu, X., Cahill, D.G., and Braun, P.V.: Colloidal metal particles as probes of nanoscale thermal transport in fluids. Phys. Rev. B **66**, 224301 (2002)
35. Huxtable, S.T., Cahill, D.G., Shenogin, S., Xue, L., Ozisik, R., Barone, P., Usrey, M., Strano, M.S., Siddons, G., Shim, M., and Keblinski, P.: Interfacial heat flow in carbon nanotube suspension. Nature Materials 2 (2003)
36. Shenogin, S,. Xu, L., Ozisik, R., Cahill, D., and Keblinski, P.: Role of thermal boundary resistance on the heat flow in carbon-nanotube composites. J. Appl. Phys. **95**, 8136–8144 (2004)
37. Domingues, G., Volz, S., Joulain, K., and Greffet, J.-J.: Heat transfer between two nanoparticles through near field interaction. Phys. Rev. Lett. **94**, 085901 (2005)
38. Ben-Abdallah, P.: Heat transfer through near-field interactions in nanofluids. Appl. Phys. Lett **89**, 113117 (2006)
39. Hashin, Z., and Shtrikman, S.: A variational approach to the theory of the effective magnetic permeability of multiphase materials. J. Appl. Phys. **33**, 3125 (1962)
40. Landauer, R.: The electrical resistance of binary metallic mixtures. J. Appl. Phys. **23**, 779–784 (1952)
41. Kim, S.H., Choi, S.R., and Kim, D.: Thermal conductivity of metal-xxide nanofluids: Particle size dependence and effect of laser irradiation. J. Heat Transfer **129**, 298–307 (2007)
42. Prasher, R., Song, D., Wang, J., and Phelan, P.: Measurements of nanofluid viscosity and its implications for thermal applications. Appl. Phys. Lett. **89**, 133108 (2006)
43. Prasher, R., Evans, W., Meakin, P., Fish, J., Phelan, P., and Keblinski, P.: Effect of aggregation on thermal conduction in colloidal nanofluids. Appl. Phys. Lett. **89**, 143119 (2006)

44. Keblinski, P., Prasher, R. and Eapen, J.: Thermal conductance of nanofluids: Is the controversy over? J. Nanopart. Res. Published online: www.springerlink.com/content/5w710jn586w38v25/ (2008)
45. Prasher, R., Phelan, P.E., and Bhattacharya, P.: Effect of aggregation kinetics on the thermal conductivity of nanoscale colloidal solutions (nanofluid). Nano Lett. **6**, 1529–1534 (2006)
46. Kwak, K., and Kim, C.: Viscosity and thermal conductivity of copper oxide nanofluid dispersed in ethylene glycol. Korea–Australian Rheology Journal **17**, 35–40 (2005)
47. Eastman, J.A., Choi, S.U.S., Li, S., Thompson, L.J., and Lee, S.: Enhanced thermal conductivity through the development of nanofluids, 3–11. Materials Research Society (MRS): Fall Meeting, Boston, USA (1997)
48. Every, A.G., Tzou, Y., Hasselmanan, D.P.H., and Raj, R.: The effect of particle size on the thermal conductivity of ZnS/diamond composites. Acta Metallurgica et Materialia **40**, 123–129 (1992)
49. Hong, T.K., Yang, H.S., and Choi, C.J.: Study of the enhanced thermal conductivity of Fe nanofluids. J. Appl. Phys. **97**, 064311 (2005)
50. Hwang, Y., Lee, J.K., Lee, C.H., Jung, Y.M., Cheong, S.I., Lee, C.G., Ku, B.C., and Jang, S.P.: Stability and thermal conductivity characteristics of nanofluids. Thermochimica Acta **455**, 70–74 (2007)
51. Lee, S., Choi, S.U.S., Li, S., and Eastman, J.A.: Measuring thermal conductivity of fluids containing oxide nanoparticles. J. Heat Transfer **121**, 280–289 (1999)
52. Li, C.H., and Peterson, G.P.: Experimental investigation of temperature and volume fraction variations on the effective thermal conductivity of nanoparticle suspensions (nanofluids). J. Appl. Phys. **99**, 084314 (2006)
53. Masuda, H., Ebata, A., Teramae, K., and Hishinuma, N.: Alteration of thermal conductivity and viscosity of liquid by dispersing ultra-fine particles (dispersion of γ-Al_2O_3, SiO_2, and TiO_2 ultra-fine particles). Netsu Bussei (Japan) **7**, 227–233 (1993)
54. Murshed, S.M.S., Leong, K.C., and Yang, C.: Determination of the effective thermal diffusivity of nanofluids by the double hot-wire technique. J. Phys. D Appl. Phys. **39**, 5316–5322 (2006)
55. Shaikh, S., Lafdi, K., and Ponnappan, R.: Thermal conductivity improvement in carbon nanoparticle doped PAO oil: An experimental study. J. Appl. Phys. **101**, 064302 (2007)
56. Wen, D., and Ding, Y.: Effective thermal conductivity of aqueous suspensions of carbon nanotubes (carbon nanotube nanofluids). J. Thermophys. Heat Transfer **18**, 481–485 (2004)
57. Wen, D., and Ding, Y.: Experimental investigation into convective heat transfer of nanofluids at the entrance region under laminar flow conditions. Int. J. Heat and Mass Transfer **47**, 5181–5188 (2004)
58. Wen, D., and Ding, Y.: Natural convective heat transfer of suspensions of titanium dioxide nanoparticles (nanofluids). IEEE Trans. Nanotech. **5**, 220–227 (2006)

Chapter 9
Nanoengineered Materials for Thermoelectric Energy Conversion

Ali Shakouri and Mona Zebarjadi

In this chapter we review recent advances in nanoengineered materials for thermoelectric energy conversion. We start by a brief overview of the fundamental interactions between heat and electricity, i.e., thermoelectric effects. A key requirement to improve the energy conversion efficiency is to increase the Seebeck coefficient (S) and the electrical conductivity (σ), while reducing the thermal conductivity (κ). Nanostructures make it possible to modify the fundamental trade-offs between the bulk material properties through the changes in the density of states and interface effects on the electron and phonon transport. We will review recent experimental and theoretical results on superlattice and quantum dot thermoelectrics, nanowires, thin-film microrefrigerators, and solid-state thermionic power generation devices. In the latter case, the latest experimental results for semimetal rare-earth nanoparticles in a III–V semiconductor matrix as well as nitride metal/semiconductor multilayers will be discussed. We will briefly describe recent developments in nonlinear thermoelectrics, as well as electrically pumped optical refrigeration and graded thermoelectric materials. It is important to note that, while the material thermoelectric figure of merit ($Z = S^2\sigma/\kappa$) is a key parameter to optimize, one has to consider the whole system in an energy conversion application, and system optimization sometimes places other constraints on the materials. We will also review challenges in the experimental characterization of thin film thermoelectric materials. Finally, we will assess the potential of some of the more exotic techniques such as thermotunneling and bipolar thermoelectric effects.

9.1 Introduction

Energy consumption in our society is increasing rapidly. A significant fraction of the generated energy is lost in the form of heat. This loss is largest in the transportation sector and in electrical power generation. Total waste heat is more than 60% of the input energy in the case of the United States. Direct thermal-to-electrical energy

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_9, © Springer-Verlag Berlin Heidelberg 2009

conversion systems that could operate at lower temperatures (100–700°C) with high efficiencies ($>$ 15–20%) will significantly expand the possibilities for waste heat recovery applications. Vining in a recent article [1] entitled *An Inconvenient Truth about Thermoelectrics* says that:

> Despite recent advances, thermoelectric energy conversion will never be as efficient as steam engines. That means thermoelectrics will remain limited to applications served poorly or not at all by existing technology.

An analysis of the potential of thermoelectrics which focuses only on efficiency values cannot be complete. It is true that thermoelectrics are not likely to replace Rankin or Stirling engines in the near future, but they could play a big role in waste heat recovery in our society. What matters is not only the efficiency, but also the cost per watt. Many groups are working on polymer and thin film solar cells. This is not because they have higher efficiency than silicon photovoltaics, but because cost is the main driving force. It is very hard to analyze the cost limits of a given technology, and in particular to make predictions about the potential changes in the future. However, this is essential in order to evaluate the potential of thermoelectrics in improving energy efficiency, and their role in the overall energy picture. In addition to the conventional use in industrial waste heat recovery and in niche cooling applications, there is a huge potential for distributed power generation in poor countries. There are many communities who cannot afford the cost of power plants and an electricity grid. However, a small amount of electricity produced by thermoelectric modules in cooking stoves or solar thermal systems could significantly improve the quality of life [2, 3]. As pointed out in Vining's article [1], solid-state thermoelectric energy conversion is already competitive with mechanical systems in small size applications.

In this chapter we do not focus on the applications. We review the basic physical principles behind solid-state thermoelectric energy conversion, as well as recent advances in nanoengineered materials and devices.

9.2 Thermoelectric Energy Conversion

Accompanying the motion of charges in conductors or semiconductors, there is also an associated energy and entropy transport. Consider a current flowing through a pair of n-type and p-type semiconductors connected in series as shown in Fig. 9.1a. The electrons in the n-type material and the holes in the p-type material all carry heat away from the top metal–semiconductor junction, which leads to cooling at the junction. This is called the Peltier effect. Conversely, if a temperature difference is maintained between the two ends of the materials as shown in Fig. 9.1b, electrons and holes with higher thermal energies will diffuse to the cold side, creating a potential difference that can be used to power an external load.

The simplified picture, which says that the difference in the electron and hole carrier signs results in different thermoelectric voltage signs, is not strictly correct.

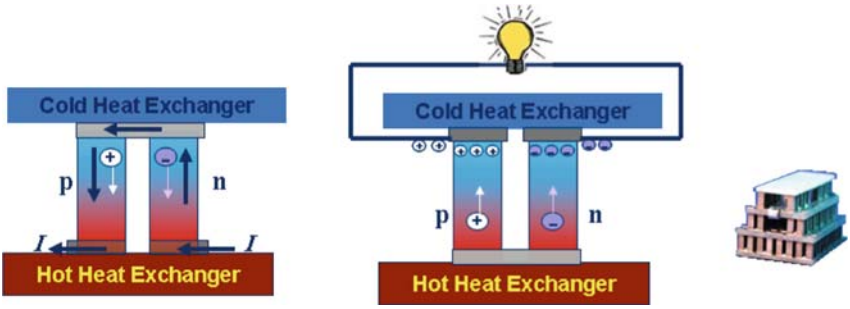

Fig. 9.1 Thermoelectric devices. *Left*: Cooler based on Peltier effect. *Center*: Power generator based on Seebeck effect. *Right*: An actual module

The asymmetry in the density of states near the Fermi level determines the sign of the Seebeck coefficient, as we will see later on. This Seebeck effect is the principle for thermocouples. For each material, the cooling effect is gauged by the Peltier coefficient Π that relates the heat carried by the charges to the electrical current through $Q = \Pi I$. The power generation is measured by the Seebeck coefficient S, which relates the voltage generated to the temperature difference through $\Delta V = S\Delta T$. The Peltier and the Seebeck coefficients are related through the Kelvin relation $\Pi = ST$. This is a consequence of Onsager's reciprocity relation.

Practical devices are made of multiple pairs of p-type and n-type semiconductor legs as shown in Fig. 9.1c. This is necessary, as thermoelectric legs require high current densities and low voltages. Putting many elements electrically in series and thermally in parallel increases the operating voltage of the module while reducing its electric current. This will minimize parasitic losses in the series electrical resistance of the wires and interconnects. The heat balance equation shows that efficient thermoelectric coolers and power generators should have a large figure of merit [4]:

$$Z = \frac{\sigma S^2}{\kappa}\,, \tag{9.1}$$

where σ is the electrical conductivity, κ the thermal conductivity, and S the Seebeck coefficient. The electrical conductivity σ enters Z through the Joule heating in the element. Naturally, the Joule heat should be minimized by increasing the electrical conductivity. The thermal conductivity κ appears in the denominator of Z because the thermoelectric elements also act as thermal insulation between the hot and cold sides. A high thermal conductivity causes too much heat leakage through heat conduction.

Because Z has units of inverse temperature, the dimensionless figure of merit ZT is often used. Shastry has recently shown that ZT, or more correctly the ratio $Z^*T/(1+Z^*T)$, is the fundamental coupling parameter between electrical charge transport and thermal energy transport by electrons in a material [5]. Z^* is the high frequency figure of merit. $Z^*T/(1+Z^*T)$ plays the same role as $C_\mathrm{p}/C_\mathrm{v} - 1$, which

is the coupling constant between sound and energy modes in anharmonic lattices or in superfluid systems. C_p and C_v are the constant pressure and constant volume heat capacities. ZT also shows up in the expression for the noise current in materials [6].

The best ZT materials are found in doped semiconductors [7]. Insulators have poor electrical conductivity. Metals have relatively low Seebeck coefficients. In addition, the thermal conductivity of a metal, which is dominated by electrons, is proportional to the electrical conductivity, as dictated by the Wiedemann–Franz law. The ratio of the electrical conductivity over electronic thermal conductivity (Lorenz number) is a function of the band structure, and it can be modified, if for instance the width of the band is reduced. For example, for the case of transport over a single energy level, the Lorenz number could go to zero [8]. It is thus hard to realize high ZT values in conventional metals. As we will see later on, thermionic current and hot electron filtering can improve the thermoelectric properties of degenerate semiconductors and metals. In semiconductors, the thermal conductivity consists of contributions from electrons (κ_e) and phonons (κ_p), with the majority contribution coming from phonons. The phonon thermal conductivity can be reduced without causing too much reduction in the electrical conductivity.

A proven approach to reducing the phonon thermal conductivity is through alloying [9]. The mass difference scattering in an alloy reduces the lattice thermal conductivity significantly without much degradation to the electrical conductivity. The traditional cooling materials are alloys of Bi_2Te_3 with Sb_2Te_3 (such as $Bi_{0.5}Sb_{1.5}Te_3$, p-type) and Bi_2Te_3 with Bi_2Se_3 (such as $Bi_2Te_{2.7}Se_{0.3}$, n-type), with a ZT at room temperature approximately equal to one [7]. A typical power generation material is the alloy of silicon and germanium, with a $ZT \approx 0.6$ at 700°C.

Figure 9.2 plots the theoretical coefficient of performance (COP) and efficiency of thermoelectric coolers and power generators for different ZT values. Also marked in the figure for comparison are other cooling and power generation technologies. Materials with $ZT \sim 1$ are not competitive against the conventional fluid-based cooling and power generation technologies. The main advantages are small form factor, flexible design (different shapes), and most importantly, no moving mechanical parts, which makes them clean and noise free. Thus, solid-state cooler and power generators have only found applications in niche areas, such as cooling of semiconductor lasers or car seat climate control systems, and power generation for deep space exploration, although the application areas have been steadily increasing.

While the search for good thermoelectrics before the 1990s was mainly limited to bulk materials, there has been extensive research in the area of artificial semiconductor structures over the last 30 years. Various means of producing ultrathin and high quality crystalline layers (such as molecular beam epitaxy and metalorganic chemical vapor deposition) have been used to alter the 'bulk' characteristics of the materials. Drastic changes are produced by altering the crystal periodicity (e.g., by depositing alternating layers of different crystals), or by altering the electron dimensionality [by confining the carriers in a plane (quantum well) or in a line (quantum wire, etc.)]. Quantum effect electronic and optoelectronic devices are used in everyday applications such as quantum well lasers in compact discs or high electron mobility transistors in cell phone base stations.

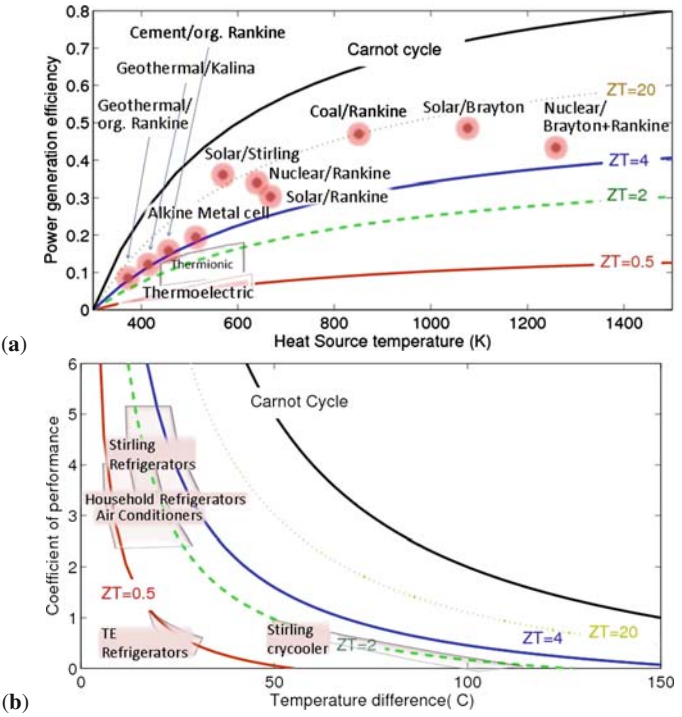

Fig. 9.2 Comparison of thermoelectric technology with other energy conversion methods for (**a**) power generation [1] and (**b**) cooling

Even though the electrical and optical properties of these artificial crystalline structures have been extensively studied, much less attention has been paid to their thermal and thermoelectric properties. Thermoelectric properties of low-dimensional structures started to attract attention in the 1990s, in parallel with renewed interest in certain advanced bulk thermoelectric materials. Some of the best known advanced bulk materials are skutterudites [10], phonon glass/electron crystal (PGEC) materials [11], and nanostructured bulk materials [12]. Research on bulk materials emphasizes the reduction of thermal conductivity. However, there are new approaches to enhance the power factor in such materials as well. Recently, Heremans et al. [13] were able to enhance the Seebeck coefficient of bulk PbTe by distorting the electronic density of states and engineering the band structure by introducing resonant thallium impurity levels in the bulk material.

Nanostructures offer the chance of improving both the electron and phonon transport through the use of quantum and classical size and interface effects. Several directions have been explored, such as quantum size effects for electrons [14, 15], thermionic emission at interfaces [16, 17], and interface scattering of phonons [18, 19]. Impressive ZT values have been reported in some low-dimensional structures [20, 21]. Some earlier publications reported ZT values in excess of 2–2.5.

However, recent detailed measurements of the thermal conductivity and thermoelectric power factor suggest that some of these numbers need to be readjusted [22].

With recent advanced scanning probe and microfabrication technologies, thermoelectric coefficients can be measured with nanometer resolution [23]. Even though there are no theoretical limits on the power factor, it has been observed experimentally that most of the enhancement in the performance of the low-dimensional materials has been due to lowering of the thermal conductivity. In 1D structures, reduction of the thermal conductivity by three orders of magnitude has been observed in single-crystalline arrays of PbTe nanowires at low temperature [24], and by 100 times in rough silicon nanowires at room temperature [25]. Shi has recently pointed out the difficulty in extracting thermal properties of individual nanowires via suspended microheater structures [26]. Additional measurements will be needed to fully characterize the thermoelectric properties of rough nanowires. In a multilayer structure, an ultralow thermal conductivity of 0.03–0.05 was measured at room temperature, six times lower than the alloy limit and only slightly above the thermal conductivity of air [27]. This exceptionally low thermal conductivity was attributed to the crystallinity of each layer and the randomness in the alignment between different layers.

Comprehensive reviews on progress in thermoelectric materials research is presented in a recently published series [10], and in the proceedings of the various International Conferences on Thermoelectrics held in recent years. In this chapter, we focus on nanoengineered materials and various techniques used to alter all three material parameters important for thermoelectric energy conversion. We also focus on thermionic emission and hot electron filtering, which can be used to improve the thermoelectric power factor (Seebeck coefficient squared times electrical conductivity).

Because of the broad scope of the work being carried out, the cited references are far from complete. Along with the review, we hope to stimulate readers by pointing out challenging, unsolved questions related to the theory, characterization, and device development of nanostructured thermoelectric materials.

9.3 Theoretical Modelling

9.3.1 Boltzmann Transport and Thermoelectric Effects

In solid-state coolers or power generators, heat is carried by charges from one place to another. In conventional materials, normal modes (quasi-particles, e.g., electrons with a given effective mass) can be defined. The current density and heat flux carried by electrons can be expressed as [28]

$$\mathbf{J}(\mathbf{r}) = \frac{1}{4\pi^3} \iiint q\mathbf{v}(\mathbf{k}) f(\mathbf{r},\mathbf{k}) \mathrm{d}^3\mathbf{k} \,, \tag{9.2}$$

$$\mathbf{J}_Q(\mathbf{r}) = \frac{1}{4\pi^3}\iiint \left[E(\mathbf{k}) - E_\mathrm{f}(\mathbf{r})\right]\mathbf{v}(\mathbf{k}) f(\mathbf{r},\mathbf{k})\mathrm{d}^3\mathbf{k}\,, \tag{9.3}$$

where q is the unit charge of each carrier, $E_\mathrm{f}(\mathbf{r})$ is the Fermi energy, $\mathbf{v}(\mathbf{k})$ is the carrier velocity, and the integration is over all possible wave vectors $\mathbf{k}$ of all the charges. The carrier probability distribution function $f(\mathbf{r},\mathbf{k})$ is governed by the Boltzmann equation, which is basically a balance showing the change in carrier distribution under external forces and scattering processes. Considering transport processes occurring much more slowly than the relaxation process, and employing the relaxation time approximation, the Boltzmann equation can be expressed as

$$\mathbf{v}\cdot\nabla_\mathbf{r} f + \frac{q\mathbf{F}}{\hbar}\cdot\nabla_\mathbf{k} f = -\frac{f(\mathbf{r},\mathbf{k}) - f_\mathrm{eq}(\mathbf{r},\mathbf{k})}{\tau(\mathbf{k})}\,, \tag{9.4}$$

where $\mathbf{F}$ is the electric field, $\tau(\mathbf{k})$ is the momentum-dependent relaxation time, $\hbar$ is the Planck constant divided by 2π, and $f_\mathrm{eq}(\mathbf{r},\mathbf{k})$ is the equilibrium distribution function for electrons (or holes), given by

$$f_\mathrm{eq}(\mathbf{r},\mathbf{k}) = \frac{1}{1+\exp\dfrac{E(\mathbf{k}) - E_\mathrm{f}(\mathbf{r})}{k_\mathrm{B}T(\mathbf{r})}}\,. \tag{9.5}$$

Here k_B is the Boltzmann constant and $T(\mathbf{r})$ is the local temperature. Under the further assumption that the local deviation from equilibrium is small, the Boltzmann equation can be linearized and its solution expressed as

$$f(\mathbf{r},\mathbf{k}) = f_\mathrm{eq}(\mathbf{r},\mathbf{k}) + \tau(\mathbf{k})\mathbf{v}\left(-\frac{\partial f_\mathrm{eq}}{\partial E}\right)\cdot\left[-\frac{E(\mathbf{k}) - E_\mathrm{f}}{T}\nabla_\mathbf{r} T + q\left(\mathbf{F} + \frac{1}{q}\nabla_\mathbf{r} E_\mathrm{f}\right)\right]\,. \tag{9.6}$$

In k-space, the distribution function at equilibrium is a Fermi sphere. When an electric field is applied, the sphere moves in the direction of the applied field and also expands (it heats up, because the electrons gain energy from the applied field) (see Fig. 9.3). Substituting (9.4) and (9.6) into (9.2) and (9.3) leads to

$$\mathbf{J}(\mathbf{r}) = q^2 L_0\left(\mathbf{F} - \frac{1}{q}\nabla_\mathbf{r} E_\mathrm{f}\right) + \frac{q}{T}L_1(-\nabla_\mathbf{r} T)\,, \tag{9.7}$$

$$\mathbf{J}_Q(\mathbf{r}) = qL_1\left(-\frac{1}{q}\nabla\Phi\right) + \frac{1}{T}L_2(-\nabla T)\,, \tag{9.8}$$

where Φ is the electrochemical potential ($-\nabla\Phi/q = \mathbf{F} + \nabla E_\mathrm{f}/q$) and the transport coefficients L_n are defined by the following integral:

$$L_n = \frac{1}{4\pi^3}\iiint \tau(\mathbf{k})v(\mathbf{k})v(\mathbf{k})\left[E(\mathbf{k}) - E_\mathrm{f}\right]^n f(\mathbf{r},\mathbf{k})\left(-\frac{\partial f_\mathrm{eq}}{\partial E}\right)\mathrm{d}^3\mathbf{k}\,. \tag{9.9}$$

From the expressions for $\mathbf{J}$ and $\mathbf{J}_Q$, various material parameters such as the electrical conductivity, the thermal conductivity due to electrons, and the Seebeck coefficient

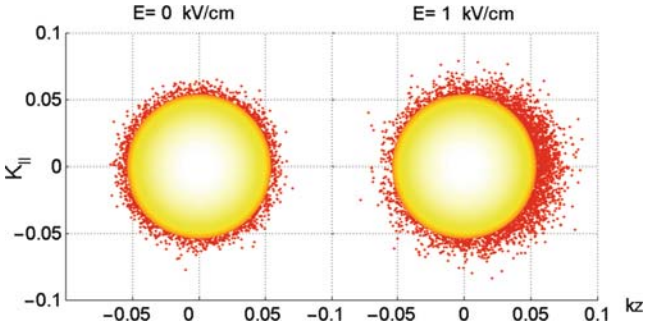

Fig. 9.3 Distribution of electrons in k-space. The *yellow sphere* has radius of k_f and is centered at the origin (equilibrium Fermi sphere) and *dots* are electrons. *Left*: Zero electric field. *Right*: Applied field of 1 kV/cm in the z direction

can be calculated. For simplicity, we assume that both the current flow and the temperature gradient are in the x direction:

$$\sigma = J_x/(-\nabla\Phi/q)\Big|_{\nabla_x T=0} = q^2 L_0 \,, \tag{9.10}$$

$$S = (-\nabla\Phi/q)/\nabla_x T\Big|_{J_x=0} = \frac{1}{qT} L_0^{-1} L_1 \,, \tag{9.11}$$

$$k_e = J_{Q_x}/\nabla_x T\Big|_{J_x=0} = -\frac{1}{T} L_1 L_0^{-1} L_1 + \frac{1}{T} L_2 \,. \tag{9.12}$$

We rewrite the expressions for the electrical conductivity and the thermopower (Seebeck coefficient) in the form of integrals over the electron energy:

$$\sigma \equiv \int \sigma(E) \left(-\frac{\partial f_{eq}}{\partial E}\right) \mathrm{d}E \,, \tag{9.13}$$

$$S \equiv \frac{k_B}{q} \frac{\int \sigma(E) \dfrac{E - E_f}{k_B T} \left(-\dfrac{\partial f_{eq}}{\partial E}\right) \mathrm{d}E}{\int \sigma(E) \left(-\dfrac{\partial f_{eq}}{\partial E}\right) \mathrm{d}E} \propto \left\langle E - E_f \right\rangle \,, \tag{9.14}$$

where we have introduced the *differential conductivity*

$$\sigma(E) \equiv q^2 \tau(E) \iint v_x^2(E, k_y, k_z) \mathrm{d}k_y \mathrm{d}k_z \approx q^2 \tau(E) \bar{v}_x^2(E) D(E) \,, \tag{9.15}$$

with $D(E)$ the density of states. $\sigma(E)$ is a measure of the contribution of electrons with energy E to the total conductivity. It is sometimes called the transport factor. The Fermi 'window' factor $(-\partial f_{eq}/\partial E)$ is a bell-shaped function centered at $E = E_f$, having a width of $\sim k_B T$. At a finite temperature, only electrons near the Fermi surface contribute to the conduction process. In this picture, the Peltier coefficient (Seebeck coefficient times absolute temperature) is the 'average' energy transported

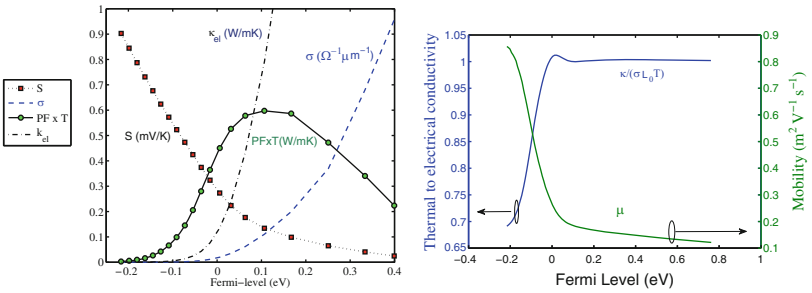

Fig. 9.4 Thermoelectric properties of GaAs. (**a**) Seebeck coefficient, electrical conductivity, electronic thermal conductivity, and power factor versus Fermi level. (**b**) Mobility and the ratio of thermal to electrical conductivity divided by temperature divided by the Lorenz number L_0

by the charge carriers. In order to achieve the best thermoelectric properties, $\sigma(E)$, within the Fermi window, should be as big as possible to increase the electrical conductivity, and at the same time, as asymmetric as possible with respect to the Fermi energy in order to enhance the thermopower. Figure 9.4 shows the calculated thermoelectric transport properties for GaAs.

9.3.2 *Theory of Thermoelectric Transport in Multilayers and Superlattices*

Electron transport is modeled by a bulk-type linear Boltzmann equation with a correction due to the quantum mechanical transmission above and below the barrier [29]. Since the optimum Fermi energy is close to barrier height and 3D states contribute significantly to electronic transport, it is also important to consider both 2D states in the wells and 3D states in the barrier. The number of electrons that participate in the thermionic emission process can be written directly as an integral in momentum space:

$$n_{\mathrm{e}}(V) = \frac{1}{L_{\mathrm{w}}}\frac{1}{\pi^2}\sum_{k_{z_i}}\int_{-\infty}^{\infty}\mathrm{d}k_x\int_{-\infty}^{\infty}\mathrm{d}k_y\Big[f(k,E_{\mathrm{f}})-f(k,E_{\mathrm{f}}-qV)\Big]T(k_{z_i},V) \tag{9.16}$$

$$+\frac{1}{2\pi^3}\left\{\frac{L_{\mathrm{w}}}{L_{\mathrm{p}}}\int_{k_{\mathrm{b}}}^{\infty}\mathrm{d}k_z\frac{\hbar^2k_z^2}{m_{\mathrm{w}}^*}\int_{-\infty}^{\infty}\mathrm{d}k_x\int_{-\infty}^{\infty}\mathrm{d}k_y\left[-\frac{\partial f(k,E_{\mathrm{f}})}{\partial E}\right]T(k_z,V)\right.$$

$$\left.+\frac{L_{\mathrm{b}}}{L_{\mathrm{p}}}\int_{0}^{\infty}\mathrm{d}k_z\frac{\hbar^2k_z^2}{m_{\mathrm{b}}^*}\int_{-\infty}^{\infty}\mathrm{d}k_x\int_{-\infty}^{\infty}\mathrm{d}k_y\left[-\frac{\partial f(k,E_{\mathrm{f}}-E_{\mathrm{b}})}{\partial E}\right]T\Big(\sqrt{k_z^2+k_{\mathrm{b}}^2},V\Big)\right\}.$$

Here V is the applied voltage over each period of the multilayer, $f(\mathbf{k},E)$ is the Fermi–Dirac distribution function, and L_{w}, L_{b}, and L_{p} are the well width, barrier

width, and superlattice period ($= L_w + L_b$), respectively. The first integral gives the contribution to the transmitted electrons from the quantized energy levels of the well (assuming quantization in the z direction). The quantum mechanical transmission probability T depends only on the V and k_{z_i} values, if we assume that the transverse momentum is conserved. In the case of non-conserved lateral momentum, we assume T to be a function of V and the total momentum k, following the argument of Meshkov [30].

The second and third integrals are the number of transmitted electrons from the energy band above the barrier located at the well and barrier regions, respectively. The latter two integrals differ in their energy reference and the effective carrier mass. For electrons in 3D states above the barrier, we have used a bulk-type Boltzmann transport equation with Fermi window factor of $-\partial f/\partial E$, and a correction accounting for the quantum mechanical transmission through the barrier. Once we have calculated the number of electrons that can move under an electric field, we can obtain the electric current by multiplying $n_e(V)$ by the electric charge and the electron drift velocity. Similarly, the entropy current (thermal current) by carriers can be calculated by adding the electron energy difference with respect to Fermi level ($E - E_f$) in the integrand of the above equation. These are approximate expressions, since we assume that all electrons have the same mobility.

In order to verify our theoretical modeling of electron transport, we first applied the theory to predict variable temperature current–voltage characteristics of the multi-quantum well structures used, e.g., in infrared detector applications [31]. A vast literature is available with experimental results for the dark current in III–V superlattices. At low temperatures, the current in the device is extremely sensitive to temperature. It can easily change by 4–8 orders of magnitude with a slight temperature increase of 50–100 K. Our theoretical curves matched experimental results that assume conservation of lateral momentum in the planar superlattices [29]. We further verified the theory by analyzing the cross-plane Seebeck coefficient in short-period InGaAs/InAlAs superlattices (lattice matched in InP) [32]. In these structures, as the doping is increased, the Seebeck coefficient exhibits non-monotonic behavior. This is quite different from bulk materials, and it is due to the formation of superlattice minibands. Theoretical curves matched well with the experimental results for the 4 samples with different dopings over a wide temperature range.

One of the shortcomings of the transport formalism presented earlier is that it does not take into account the change in electron effective mass (group velocity), which could be important in narrow-band superlattice structures. Bian et al. have developed a self-consistent solution to the Schrödinger equation in the superlattice together with the Poisson equation [33]. This was needed to model band bending, which results from charge transfer between the barrier and the well regions. Subsequently, the superlattice dispersion equation [energy–momentum relation $E(\mathbf{k})$] and a modified differential conductivity were used to calculate the relevant transport parameters (electrical conductivity, Seebeck coefficient, and electronic contribution to the thermal conductivity). The simulation results matched well with the cross-plane Seebeck coefficient in 20 nm InGaAs/ 10 nm InGaAlAs superlattices [33]. The barrier height was 0.2 eV and the layers were lattice-matched to the InP substrate. The

number of charge carriers in InGaAs wells was 1×10^{19} cm^{-3}, which came from the silicon dopants as well as 0.03% ErAs nanoparticles. It is interesting to note that the calculation of the electronic contribution to the thermal conductivity showed that the superlattice structure can cause changes in the Lorenz number by as much as a factor of 2 compared to the bulk material.

9.3.3 Monte Carlo Simulation of Electron Transport in Thermoelectric Layers

Bian and Shakouri have developed a Monte Carlo program in order to investigate how non-planar barriers can affect electron transport and evaluate non-conservation of lateral momentum [34]. This program can calculate the average number of electrons transmitted above an arbitrary shaped two-dimensional potential barrier. The average transport energy of the transmitted electrons, i.e., the Seebeck coefficient, can also be calculated. Simulation results show that non-planar barriers do indeed have larger thermoelectric power factors compared to planar ones.

Zebarjadi et al. has developed the first complete Monte Carlo program [35] to simulate thermoelectric transport in the III–V family of materials. The code is three-dimensional in both **k** and **r** space, with non-parabolic multivalley band structure. The scattering mechanisms included are: ionized and neutral impurities, intra-valley polar optical phonons, acoustic phonons, and inter/intra-valley non-polar optical phonons. The Pauli exclusion principle is critical, as optimum thermoelectric materials are nearly degenerate. Direct estimation of the Pauli exclusion principle using iterative calculation of the local electron density is computationally very expensive. Instead, the Pauli principle was enforced after each scattering process, supposing a shifted Fermi sphere as the local electronic distribution. For each valley, the electronic temperature is defined locally as follows:

$$f_v(E, E_\mathrm{f}^v, T_\mathrm{e}^v) = \left[1 + \exp \frac{E_v\left(|\mathbf{k} - \mathbf{k}_\mathrm{d}^v(\mathbf{r})|\right) - E_\mathrm{f}^v(\mathbf{r})}{k_\mathrm{B} T_\mathrm{e}^v(\mathbf{r})}\right]^{-1}, \tag{9.17}$$

$$T_\mathrm{e}^v(\mathbf{r}) = \frac{2}{3k_\mathrm{B}}\left[\left\langle E_v\left(\mathbf{k} - \mathbf{k}_\mathrm{d}^v(\mathbf{r})\right)\right\rangle - \left\langle E_v(\mathbf{r})\right\rangle_0\right] + T\,, \tag{9.18}$$

where $\left\langle E_v(\mathbf{r})\right\rangle_0$ is the local average energy of electrons in equilibrium at zero electric field, $\mathbf{k}_\mathrm{d}^v(\mathbf{r})$ is the local drift wave vector, which is the average wave vector of all the particles at position **r** and in the valley v, T is the lattice temperature, and E_f is the quasi-Fermi level [36]. The resulting distribution functions inside a bulk material are shown in Fig. 9.5. The program was used to simulate both bulk and multilayer (inhomogeneous systems). Electrons are injected through the contact–electrode junction using the Fermi distribution of the same material as the contact layer.

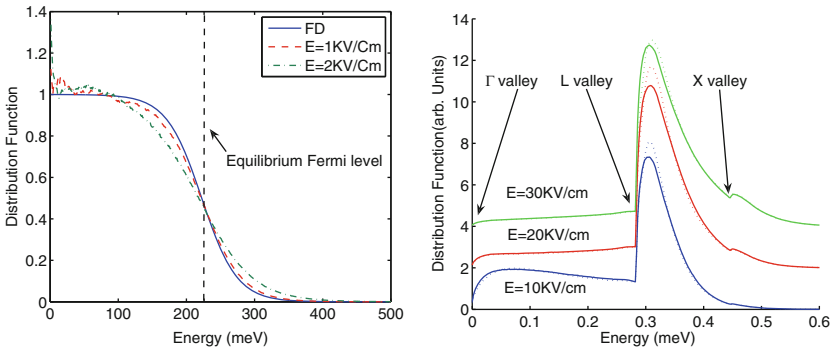

Fig. 9.5 Monte Carlo simulation. Distribution function of bulk GaAs under applied electric field. (**a**) Low field. (**b**) High field. For greater clarity, the three graphs are shifted upwards by two units, and different valley minima are shown with an *arrow*

Most previous work has considered the Peltier effect as a localized energy exchange that happens only at the interface. Recent Monte Carlo simulations have shown that Peltier cooling and heating happen mostly inside the highly doped or metal contacts (and not inside the semiconductor, which is the main thermoelectric material) (see Sect. 9.12). The size of the cooling/heating region can be ~ 0.2–0.4 μm. Since thermoelectric (TE) energy exchange happens in the contact layers, increasing the thermal interface resistance between the metal and semiconductor can improve the cooling performance of short-leg TE coolers. By studying the spatial distribution of thermoelectric heat exchange, we can engineer the lattice thermal conductivity near the interface in order to maximize the TE device performance. In the case of very short barriers (ballistic transport in superlattices), the spatial distribution of thermionic energy exchange is also important in optimizing the superlattice design (well and barrier thicknesses, etc.).

9.3.4 Non-Equilibrium Green Function for Thermoelectric Transport

The Monte Carlo technique is powerful enough to calculate the thermoelectric properties of homogeneous and inhomogeneous materials in the classical (point particle) regime. As we have seen, in certain quantum systems, when electrons remain coherent e.g., over several superlattice periods, one can use the modified Boltzmann approach, taking into account the electronic mini-bands to estimate thermoelectric properties. However, when both quantum mechanical wave properties and scattering are important, one has to solve the Schrödinger equation coupled to reservoirs. This is necessary, for example, to calculate the thermoelectric properties of individual molecules. The first calculations were done by Paulsson and Datta [37]. Recently, other authors have expanded the theory [38]. The first experimental results on the

thermoelectric properties of single-molecule junctions were presented by Reddy et al. [39], and the results are quite consistent with the calculations presented by Paulson and Datta [37].

9.3.5 Phonon Transport

The thermal conductivity of phonons is also often modeled by the Boltzmann equation under the relaxation time approximation:

$$k_{\mathrm{p}} = \frac{1}{3}\sum \int C(\omega) v_{\mathrm{p}}(\omega) \Lambda(\omega) \mathrm{d}\omega \,, \tag{9.19}$$

where C is the specific heat of phonons at frequency ω, v_{p} the phonon group velocity, and Λ the phonon mean free path.

There are several approaches for extending the Boltzmann transport equation to low dimensions. One approach is to treat the low-dimensional structures effectively as a bulk material. This approach includes modifications of acoustic phonon dispersion and group velocities due to phonon confinement [40] and appropriate boundary conditions to describe partially diffusive phonon scattering at the surfaces [41]. Another approach uses the molecular dynamics method. It allows accurate calculation of phonon dispersion and thermal conductivity of structures with a few atomic layers, but it cannot include a variety of quantum effects, it requires knowledge of the interatomic potential, and it is limited by the computation time. This approach is needed if the size of the nanostructure is smaller than the phonon mean free path [42]. The third possibility is to solve the BTE by the Monte Carlo technique, which is a semiclassical statistical method based on simulating an ensemble of particles. Although MC simulation has been widely used in electron transport, it has not been very popular in phonon transport. The main difficulty is to include the phonon–phonon interaction, which will affect the distribution function [43, 44]. Finally at the quantum level, when transport is coherent, the Landauer formalism for thermal transport has been widely used [45].

The biggest difference between electron and phonon transport is that phonons obey Bose–Einstein statistics and electrons Fermi–Dirac statistics. Without the Pauli exclusion principle, one has to include all phonon modes in the calculation of heat transport. In the case of electrons, only electrons near the Fermi surface contribute to transport.

The formulation for thermoelectric properties leads to the following possibilities to increase ZT and thus the energy conversion efficiency of devices made of nanostructures in the linear transport regime:

1. Interfaces and boundaries of nanostructures impose constraints on the electron and phonon waves, which lead to a change in their energy states and correspondingly, their density of states and group velocity.

2. The symmetry of the differential conductivity with respect to the Fermi level can be controlled using quantum size effects and classical interface effects (as in thermionic emission).
3. The phonon thermal conductivity can be reduced through interface or nanoparticle scattering and through the alteration of the phonon spectrum in low-dimensional structures.

We will see in Sects. 9.13 and 9.14 that it is possible to go beyond the linear transport regime and ZT optimization, and enhance the thermoelectric energy conversion using nonlinear effects or, e.g., by coupling electrons, phonons, and photons.

9.3.6 Thermoelectric Transport in Strongly Correlated Systems

In the case of correlated electronic systems such as sodium cobalt oxide ($NaCo_2O_4$), strong electron–electron interactions make the band picture inaccurate. As electrons move from one site to another, strong Coulomb interaction can change the energy levels depending on the occupation number in each state. In such cases, a description based on independent particles in momentum space (k-space) is incorrect. However, it is easy to develop a rate equation in real space. One can include the electron hopping probability between neighboring sites and take into account electron–electron interactions. A more rigorous approach is to solve the Hamiltonian for the system. This is the basis of the Heikes formula, which is valid at the high temperature limit [46]. Terasaki et al. were the first to measure high thermoelectric power factors for $NaCo_2O_4$ in 1997 [47]. An unusually high Seebeck coefficient for a material with high electronic conductivity resulted in power factors approaching that of Bi_2Te_3, the best commercial room temperature thermoelectric material. Unfortunately, thermal conductivity is high, so the overall thermoelectric figure of merit is limited.

Several groups have tried to explain the high Seebeck coefficient of $NaCo_2O_4$ or other strongly correlated oxide systems. Unfortunately, a unified picture has not yet emerged. Singh and Kasinathan suggest that conventional Boltzmann transport calculations can predict the measured values of the Seebeck coefficient and its temperature dependence quite accurately [48]. On the other hand, Koshibae et al. have used an argument based on the spin degree of freedom in order to explain the large Seebeck coefficient [49]. They argue that, as the electrons hop from one site to another, the spin degree of freedom is different at Co^{3+} and Co^{4+} sites. The change in the spin degree of freedom affects the amount of entropy carried by electrons. Thus the spin degree of freedom should give a contribution to the Seebeck coefficient. Following Heikes' formula, this contribution is given by $k_B/e \ln(g_3/g_4)$, where g_3 and g_4 are the degeneracies of the states in the Co^{3+} and Co^{4+} sites, respectively. This spin Seebeck coefficient of 154 or 214 μV/K (depending on the low or high spin states) is the value reached at high temperatures, and it compares well with the measured values of ~ 100 μV/K at 300 K. Recently, Peterson et al. have done detailed calculations of the thermoelectric properties of strongly correlated systems [50].

Their analysis highlights the importance of the frustrated triangular lattice in determining the electronic transport parameters. This is also evident in the anomalous Hall coefficients for these systems. The frustrated lattice increases the degeneracy of the ground state and thus the spin degree of freedom.

Since Terasaki's discovery in 1997, a lot of effort has gone into optimizing the thermoelectric properties of oxide materials. In addition to the intriguing new physics, one of the main driving forces is the fact that these oxide materials are supposed to be quite stable at high temperatures, and appropriate for waste heat recovery applications. For a summary of the Japanese program (CREST), one should consult papers by Ohta et al. [51]. Recently, Scullin et al. at the University of California, Berkeley and Berkeley Labs have synthesized many epitaxial oxide thin films. There is a lot of potential to modify the electronic transport via heterostructures and reduce the lattice thermal conductivity by nanostructuring [52]. A key difficulty seems to be in the characterization of the epitaxial films, as the oxygen could diffuse from or into the substrate, making it conductive and thus affecting the thin-film measurements.

9.3.7 Wave or Particle Picture for Electrons and Phonons?

The transport of electrons and phonons in nanostructures is affected by the presence of the interfaces and surfaces. Since electrons and phonons have both wave and particle characteristics, the transport can fall into two different regimes: totally coherent transport, in which electrons or phonons must be treated as waves, and totally incoherent transport in which either or both of them can be treated as particles. There is, of course, the intermediate regime where transport is partially coherent, an area that has not been studied extensively. Whether a group of carriers are coherent or incoherent depends on the strength of phase-destroying scattering events (such as internal or diffuse interface scattering). In a nanostructure with no phase-destroying scattering events, a monochromatic wave can experience many coherent scatterings while preserving the phase. The coherent superposition of the incoming and scattered waves leads to the formation of new energy bands for electrons and/or phonons (i.e., in a superlattice). This changes the number of available states per unit energy (i.e., the density of states), which has a profound effect on the electrical, optical, magnetic, and thermal properties of the material. On the other hand, if there is strong internal scattering (which can be judged from the momentum relaxation time) or if the interface scattering is not phase-preserving (e.g., when due to diffuse scattering), no new energy bands form. In this regime nanostructures are still able to modify the thermoelectric properties of the material, e.g., by the selective scattering of phonons with respect to electrons (i.e., reduced lattice thermal conductivity without much reduction in the electrical conductivity). Another possibility is the selective scattering of cold (low energy) electrons, which can enhance the Seebeck coefficient. This is sometimes called thermionic emission.

In the following, we focus on two main areas where nanostructures could enhance thermoelectric energy conversion (low dimensionality and thermionic

emission). Before we describe the detailed role of nanostructures in these two transport regimes, let us consider the trade-off between the electrical conductivity and the Seebeck coefficient.

9.3.8 Why Is There a Trade-off Between Electrical Conductivity and Seebeck Coefficient?

The fact that there is a trade-off between electrical conductivity and the Seebeck coefficient, and that we cannot keep increasing the number of free carriers and get higher and higher power factors, is an intriguing effect which has not been discussed in detail in the literature [53–55]. This trade-off can be explained easily if we consider the concept of differential conductivity, introduced earlier on (see Sect. 9.3.1). In a degenerate semiconductor, when the Fermi energy is close to the band edge (bottom of the conduction band or top of the valence band), the density of states versus energy curve is asymmetric with respect to the Fermi level (see Fig. 9.6). This means that, e.g., for the n-type material, there are more states available for transport above the Fermi energy than below it. As we increase the doping in the material, the Fermi energy moves deeper into the band and the differential conductivity becomes more symmetric with respect to the Fermi energy. This is due to the fact that the density of states has a square-root dependence on energy for any typical 3D single band crystal. This can be explained by geometric considerations. Momentum is the main quantum number describing electrons in a crystal. The density of states is just a count of the number of electrons that occupy a given energy state. Since energy and momentum are related by a quadratic equation within the effective mass approximation, the number of states at a given energy scales as the surface area of the Fermi sphere in the momentum space. So in 3D materials, this surface (e.g., DOS) is proportional to the square root of the electron energy. It thus seems obvious that going to lower dimensional semiconductors can inherently improve the thermoelectric power factor by creating sharp features in the electronic DOS. Another possibility is to use an appropriate hot electron filter (potential barrier) that selectively scatters cold electrons. Here, in the near-linear transport regime, hot electrons denote carriers that contribute to electrical conduction with energies higher than the Fermi level, while cold electrons have energies lower than the Fermi level. This nomenclature differs from the one used in device physics, where hot carriers are typically non-equilibrium populations which can be built up under high electric fields.

One should note in Fig. 9.6 that, once non-parabolicity or energy-dependent effective mass is included, the density of states bears an almost linear relationship with energy deep in the band. As the doping increases, the symmetry of the DOS is constant. However, the Seebeck coefficient still keeps decreasing. This is because the denominator of equation (9.14) is proportional to the electrical conductivity. As the doping in the material and the conductivity increase, we need a larger asymmetry in the DOS if we want to keep the Seebeck coefficient high.

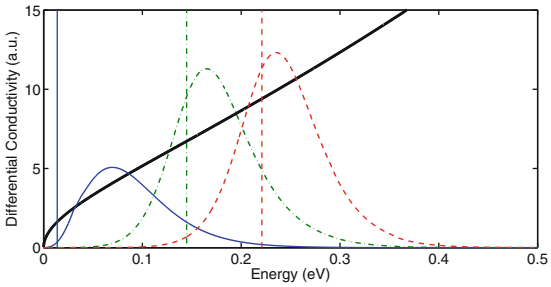

Fig. 9.6 Differential conductivity for three different Fermi levels vs. energy. The corresponding Fermi levels are plotted for each curve. The density of states is also plotted with a *black solid line*. These plots are for GaAs with non-parabolic band structure

9.3.9 Low-Dimensional Thermoelectrics

In 1993, the outstanding pioneering work of Hicks and Dresselhaus renewed interest in thermoelectrics, becoming the inspiration for most of the recent developments in the field [14]. Dresselhaus et al. showed that electrons in low-dimensional semiconductors such as quantum wells and wires have an improved thermoelectric power factor (Seebeck coefficient squared times electrical conductivity) and $ZT > 2$–3 can be achieved (see Fig. 9.7). This is due to the fact that electron motion perpendicular to the potential barrier is quantized, creating sharp features in the electronic density of states [56]. Figure 9.8 illustrates the density of states (DOS) of electrons in the bulk InSb material, as well as quantum wells and quantum wires with thickness or diameter 4 nm, respectively. Figure 9.8 also indicates that the Seebeck coefficient is large when the average electron energy is far from the Fermi level. Experimental results for transport in PbTe/PbEuTe and Si/SiGe quantum well systems indicated an increased value of ZT inside the quantum well [57, 58]. However, there remain some questions regarding the role of band bending and true quantum confinement in the early experiments with PbTe superlattices.

Quantum confinement changes the energy of the band edge of the semiconductor. Near this band edge, some sharp features are created in the DOS. One can use these sharp features to increase the asymmetry between hot and cold electron transport, and thus obtain a large average transport energy and a large number of carriers moving in the material, i.e., a large Seebeck coefficient squared times electrical conductivity. One should note that the sharp features in the density of states of quantum wells and wires increase the Seebeck coefficient at the optimum doping only modestly (by 20–50% see [59]). The order of magnitude improvement in thermoelectric power factor predicted in the literature comes mainly from the increase in the effective 3D electrical conductivity when 2D and 1D conductivity results are normalized by the width or cross-section of the wells or wires. This requires minimal surface scattering and highly dense and fully localized electrons in an array of low-dimensional structures. This is probably the reason why the enhancements in

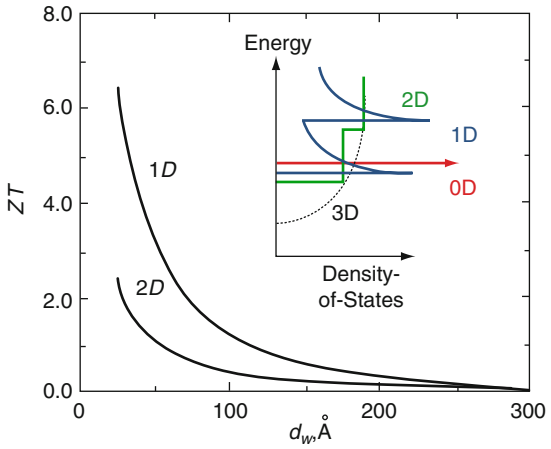

Fig. 9.7 Thermoelectric figure of merit ZT of Bi quantum wells and quantum wires as a function of dimension [14]. *Inset*: Density of states for 3D, 2D, 1D, and 0D conductors

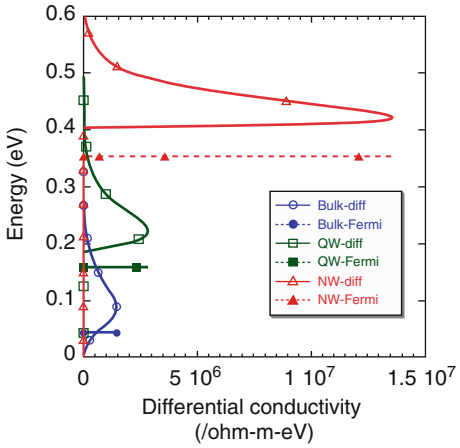

Fig. 9.8 Differential conductivity of typical bulk, quantum well, and quantum wire materials versus electron energy (infinite barriers are assumed). *Dashed lines* show the respective optimum Fermi levels for highest thermoelectric power factor. The average energy of the moving electrons with respect to the Fermi level is a definition of the Peltier coefficient. (Courtesy of Professor Z. Bian)

the thermoelectric power factor of the whole superlattices or nanowire composites have not yet been observed. However, the pioneering work of Hicks and Dresselhaus on low-dimensional thermoelectrics has been an inspiration to go beyond the trade-offs in bulk materials, and to use nanostructures to engineer the thermoelectric properties of materials.

The recent breakthroughs in materials with $ZT > 1$ (Venkatasubramanian et al. [60], Harman et al. [61], or Hsu et al. [62]) have mainly benefited from reduced phonon thermal conductivity [19], with power factors similar to the existing

state-of-the-art material. Recently, there have been some doubts about the validity of Harman's claims [22]. There are three reasons why it is hard to improve the thermoelectric power factor of quantum well materials [63–65]. First, we live in a 3D world and any quantum well structure should be embedded in barriers. These barriers are electrically inactive, but they add to thermal heat loss between the hot and the cold junctions. This can reduce the performance significantly [66]. One cannot make the barrier too thin, since the tunneling between adjacent quantum wells will broaden energy levels and reduce the improvement due to the density of states. The second reason is that the sharp features in the density of states of low-dimensional nanostructures disappear quickly as soon as there is size non-uniformity in the material. (Even though this makes sense intuitively, there are no detailed calculations of the effect of size non-uniformity on low-dimensional thermoelectric properties.) The third reason concerns interface scattering of electrons in narrow quantum structures. A natural extension of quantum wells and superlattices is to quantum wires [67,68]. Theoretical studies predict a large enhancement of ZT inside quantum wires due to additional electron confinement (see Fig. 9.7). Experimentally, different quantum wire deposition methods have been explored [69–75]. However, so far, there have been no experimental results indicating any significant enhancement of the thermoelectric power factor in individual quantum wires. In the case of nanowire arrays, the whole structure has been embedded in an alumina template or in a polymer. The difficulty in ensuring good electrical contact to all wires in an array, and quantum wire size variations have so far impeded the quantitative characterization of low-dimensional properties.

Quantum dot structures have been proposed as the 0D extension of the low-dimensional thermoelectrics [61]. In a theoretical study by Mahan and Sofo [76], it was suggested that the best thermoelectric materials will have a delta function density of states. Quantum dots fit ideally into such a picture. A single quantum dot, however, is not of much interest for building into useful thermoelectric devices (but may be of interest to create localized cooling on the nanoscale). One has to use an array of quantum dots. Recently, such arrays have been investigated theoretically and experimentally as potentially good thermoelectrics. Cai and Mahan [77] developed a dynamical mean field theory to calculate the electrical properties of a crystalline array of quantum dots. They suggest that such arrays may have high Seebeck coefficients at low temperatures.

However, there is a fundamental difference between quantum dots and quantum wires/wells. The original theory developed by Dresselhaus et al. [67] does not rigorously apply to quantum dots. The enhanced power factor in quantum confined 2D and 1D structures happens in the direction perpendicular to the confinement. Thus we benefit from sharp features in the density of states, but we can still use the free electron approximation with an effective mass in the direction in which the electric field is applied and heat is transported. However, in the case of a matrix of quantum dots, electrons have to move between the dots in order to transfer heat from one location to another. If the electronic bands in the dots are very narrow, then the electrons are highly confined, and it is not easy to take them out of the dots. On the other hand, it is easy to take electrons out of shallow energy barrier quantum dots, but at

the same time the density of electronic states in the dot will have broad features. Adding linkers between dots might solve this problem. It seems that there should be an optimum quantum dot size and and an optimum coupling between dots that gives the highest power factor. However, this has not been fully clarified. Some of the difficulty is related to the modeling that requires quantum mechanical confinement of electrons as well as scattering mechanisms as electrons move between the dots. It seems that the recent non-equilibrium Green's function formalisms could be ideal for modeling such systems [37, 38].

9.4 Thermionic Energy Conversion

9.4.1 Vacuum Thermionic Energy Conversion

Thermionic energy conversion is based on the idea that a high work function cathode in contact with a heat source will emit electrons [78]. These electrons are absorbed by a cold, low work function anode separated by a vacuum gap. They can flow back to the cathode through an external load where they do useful work. A vacuum is the best electrical conductor (electrons have no collisions with energy losses in the gap) and the worst thermal conductor, since there are no atoms to transmit random vibrations and heat is only transported via radiation. Practical vacuum thermionic generators are, however, limited by the work function of available metals or other materials that are used for cathodes and anodes. Another important limitation is caused by the space charge effect. The presence of charged electrons in the otherwise neutral space between the cathode and anode will create an extra potential barrier between the cathode and anode, which reduces the thermionic current. The materials currently used for cathodes have work functions > 0.7 eV, which limits the generator applications to high temperatures > 500 K. Mahan [79, 80] has proposed these vacuum diodes for thermionic refrigeration. Basically, the same vacuum diodes that are used for the generators will work as a cooler on the cathode side and a heat pump on the anode side under an applied bias. Mahan predicted efficiencies of over 80% of the Carnot value, but once again these refrigerators will only work at high temperatures (> 500 K).

9.4.2 Nanometer Gaps and Thermotunneling

There has been recent research to make efficient thermionic refrigerators at room temperature with the use of nanometer thick vacuum gaps [81]. This is sometimes called thermotunneling. This idea is based on the fact that electron tunneling decreases exponentially as a function of barrier thickness. Use of < 5–10 nm barriers will allow conventional metal electrodes to achieve appreciable emission currents

($\sim$ 100 A/cm^2) and cooling power densities ($\sim$ 100 W/cm^2) at room temperature. There have been detailed theoretical studies and some preliminary experimental results [82]. However, the measured cooling is very small (0.3 mK). The main difficulty to achieve substantial cooling is to produce uniform nanometer-size vacuum gaps over large areas. In addition, this narrow gap should be maintained as the temperature difference develops and the cathode and anode undergo thermal expansion and mechanical stress.

Recently, Gerstenmaier and Wachutka [83] have analyzed thermionic energy converters in the micro- and nanometer gap range. Their comprehensive theory included backward currents from the collector electrode, losses due to thermal radiation and Ohmic resistance in the electrodes, tunneling through the gap, image forces, and space charge effects. They showed that the efficiency of nanometer gap thermionic converters could be much higher than the efficiency of thermoelectric devices for operating temperatures above 800 K (assuming work functions of 1 eV). Gerstenmaier and Wachutka's analysis shows that nanometer gaps do not really remove the requirement for low work function emitters and collectors for a vacuum thermionic energy conversion device to work at low temperatures. It was shown that, in order to attain high efficiencies at low temperatures (300–500 K), work functions of 0.3–0.5 eV are necessary, even with nanometer gaps. Unfortunately, such low work functions have not yet been achieved.

9.4.3 Inverse Nottingham Effect and Carbon Nanotube Emitters

In another approach, sometimes called the inverse Nottingham effect, resonant tunneling at an appropriately engineered cathode band structure has been proposed to enhance vacuum emission currents in a narrow energy range [84, 85]. There have been no experimental demonstrations as yet [86]. Use of enhanced field emissions at nanostructured surfaces, such as carbon nanotubes or sharp tips, has also been investigated theoretically and experimentally [87,88]. While significantly increased vacuum currents have been obtained [89, 90], there are no experimental results on thermionic energy conversion. An important problem with carbon nanotube field emitters is that we do not yet have direct control of the nanotube chirality, i.e., its electrical conductivity, since both metallic and semiconducting nanotubes are grown at the same time. It is also important to note that the 'selective' emission of hot electrons (energies higher than the Fermi level) compared to cold electrons (energies lower than the Fermi level) is necessary in order to achieve energy conversion. Since at room temperature the energy distribution of electrons inside the Fermi window is on the order of 25–50 meV, a precise engineering of tunneling is necessary to achieve appreciable energy conversion. Recently, Koeck et al. have demonstrated vacuum power generation with a nanostructured nitrogen-doped diamond emitter, separated by a $\sim$ 80 μm gap from the collector. At a cathode temperature of 825°C, substantially below conventional vacuum thermionic modules, 120 mV thermovoltage was generated [91–93]. It is anticipated that vacuum thermionic emitters could

be useful for high temperature power generation. However, applications for energy conversion at low temperatures will probably not be available any time soon.

9.4.4 Single Barrier Solid-State Thermionic Energy Conversion

In the early to mid-1990s, several groups pointed out the advantage of electron energy filters in bulk thermoelectric materials [94–96]. This followed the pioneering work of Moyzhes [97]. However, these papers were not widely referenced. To overcome the limitations of vacuum thermionics at lower temperatures, thermionic emission cooling in heterostructures was proposed by Shakouri and Bowers [16, 98]. In these structures, a potential barrier is used for the selective emission of hot electrons and the evaporative cooling of the electron gas. The heterostructure integrated thermionic (HIT) cooler can be based on a single barrier or a multibarrier structure (see Fig. 9.9). In a single barrier structure in the ballistic transport regime, which is strongly nonlinear, the electric current is dominated by the supply of electrons in the cathode layer, and large cooling power densities exceeding kW/cm^2 can be achieved [99, 100]. In this design, it is necessary to use a barrier several microns thick with an optimum barrier height at the cathode, on the order of the thermal energy $k_B T$ of the electrons, where k_B is the Boltzmann constant. A large barrier height at the anode is also needed to reduce the reverse current [100]. This large barrier height will significantly increase the forward current and the cooling power density.

A few single barrier InGaAs/InGaAsP/InGaAs thin film devices, lattice matched to an InP substrate, have been fabricated and characterized [101]. The InGaAsP barrier ($\lambda_{gap} \approx 1.3$ μm) was one micron thick and ~ 100 meV high. Even though cooling by 1°C and cooling power density exceeding 50 W/cm^2 were achieved [102], it was not possible to increase the bias current substantially and benefit fully from the large thermionic emission cooling. This is due to a non-ideal metal–semiconductor

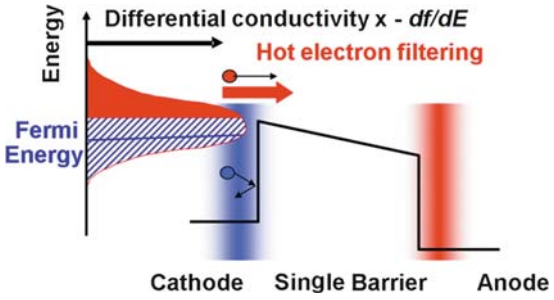

Fig. 9.9 Band diagram of a single barrier heterostructure thermionic energy converter. Selective emission of hot electrons can produce an electrical voltage under a temperature gradient. In the case of ballistic transport across the barrier, the device is in the nonlinear transport regime. If the barrier is made of a thick multibarrier or superlattice, under small biases, one can define an effective electrical conductivity and Seebeck coefficient by treating this as an effective medium

contact resistance and Joule heating in the substrate. The single barrier HIT device in a nonlinear transport regime was not anticipated to have an improved energy conversion efficiency. Electrons that are ballistically emitted release all their excess energy in the anode and produce significant heating. In general, in order to approach the Carnot limit, the transport should be quasi-reversible and near equilibrium. The main motivation of the original study was to achieve temperature stabilization of optoelectronic devices with monolithic structures [102, 103]. Recent studies have shown that there is potential benefit from *nonlinear Peltier effects* in single barrier structures for cryogenic cooling applications [104]. Basically, ballistic transport in the barrier will increase the average electron transport energy (i.e., the Peltier coefficient). This bias-dependent Peltier coefficient is strongest in low-doped semiconductors where the electron heat capacity is small. It has been shown, e.g., that the Peltier coefficient of InGaAs barriers with a doping of 5×10^{16} cm^{-3} could be doubled with a current density of ~ 150 kA/cm^2. The current density is high but common for thin film devices. As we shall see in Sect. 9.13, the nonlinear Peltier effect can increase the maximum cooling of HIT coolers by a factor of seven at cryogenic temperatures.

9.4.5 Multilayer Solid-State Thermionic Energy Conversion

For a multibarrier structure at small biases, the transport is linear and one can define an effective Seebeck coefficient and electrical conductivity. Mahan and Woods [80] were the first to linearize the conventional Richardson equation for the thermionic emission current in a multibarrier device. Their initial calculations suggested that multibarrier structures could have efficiencies twice as large as conventional thermoelectrics [80]. However, more detailed analysis by Radtke et al. [105], Ulrich et al. [106], and Mahan and Vining [107] showed that the linearized Richardson equation does not produce electronic efficiencies higher than thermoelectrics, and it was claimed that the only benefit of the multilayer structure was in reducing the lattice thermal conductivity [108]. These calculations do not give the full potential of multibarrier devices, since the focus was only on small barrier heights (conduction band discontinuity on the order of the thermal energy), and also because the authors used the linearized version of the Richardson equation, which is not a good approximation when the Fermi energy is near the barrier height. For a more accurate analysis of electron transport perpendicular to superlattice layers, a modified Boltzmann transport equation was proposed that takes into account the quantum mechanical transmission through barriers [109] (see Sect. 9.3.2 on theoretical modeling).

The motivation to work on *metal*-based superlattices and embedded nanoparticles was inspired by the theoretical calculation of Vashaee and Shakouri [109] who predicted possible values of $ZT > 5$ for optimized structures, even with a lattice contribution to thermal conductivity on the order of 1 W/mK. The main idea is that *in a thermoelectric energy conversion device, work is extracted from the random thermal motion of electrons*, so in principle we would like to have *as many*

electrons as possible in the material. However, highly degenerate semiconductors and metals are not good bulk thermoelectric materials due to their low Seebeck coefficient. It was shown earlier that highly degenerate semiconductors and metallic structures could have high thermoelectric power factors (Seebeck coefficient squared times electrical conductivity) if there is an appropriate *hot electron filter* (*potential barrier*) that selectively scatters cold electrons. In fact, highly efficient *tall* barrier metallic superlattices were first suggested in 1998 [53]. However, detailed modeling of electron transport in these structures revealed the importance of *lateral momentum non-conservation* [109], as described in the next section.

9.4.6 Conservation of Transverse Momentum in Thermionic Emission

A judicious choice of potential barriers in a highly doped semiconductors or metals can increase the asymmetry between hot and cold electron transport, thereby overcoming the conventional trade-off between electrical conductivity and the Seebeck coefficient (see Fig. 9.10a) [110]. However, the simplistic picture in the energy space is misleading. One may think that all hot electrons with energies larger than the barrier height are transported above the barrier. However, if we look at electronic states in the momentum space (Fig. 9.10b), we see that, with planar barriers, only electrons with kinetic energy in the direction perpendicular to the barrier higher than the threshold value are emitted (e.g., volume V_1 in Fig. 9.10b) [8, 111]. There are many hot electrons that have large transverse momentum. They cannot go above the barrier layer. The basic idea is that planar superlattices are *momentum filters* and not *energy filters* [112]. In an analogy with optics, we can say that these hot electrons have total internal reflection at the barrier interface, and they cannot be emitted (see Fig. 9.10c).

The conservation of transverse momentum is due to the symmetry of the system (translation invariance in the direction perpendicular to the barrier layers). Using non-planar barriers or scattering centers, one can break this symmetry [113]. The key requirement is to break the symmetry without a significant reduction in the electron mean free path (electron mobility) in the structure. Thus it is important to have a low defect density and a high crystallinity near the interface. This could be achieved with, e.g., embedded nanoparticles [114]. It is important to note that the role of nanoparticles as hot electron filters is quite different from what happens in low-dimensional thermoelectrics. Discrete energy states inside the quantum dot are not directly used. Quantum dots act as three-dimensional scattering centers and energy filters for electrons moving in the material. It is interesting to note that, if there is transverse momentum conservation, not only is the number of emitted electrons significantly reduced, but in addition, the energy filtering is not abrupt even with thick barriers [111]. Gradual selection of hot electrons results in low electronic efficiency of the structure.

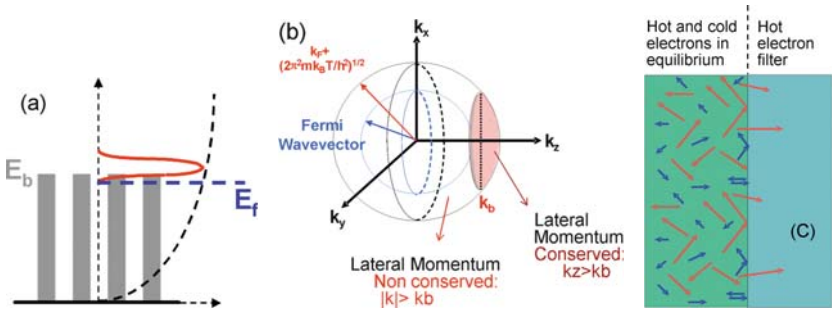

Fig. 9.10 *Left*: Schematic showing the density of states in the conduction band when the Fermi energy is deep in the band. The energy diagram of the multibarriers versus distance is superimposed to show the selective emission of hot electrons. *Center*: Representation of electronic states in momentum space when the Fermi energy is deep inside a band (Fermi sphere). *Right*: When lateral momentum is conserved, only electrons with large enough kinetic energy in the direction perpendicular to the barriers are transmitted. However, when the lateral momentum is not conserved, the number of emitted electrons increases substantially

Recently, Wang and Mingo have studied the role of rough barriers and lateral momentum non-conservation in InGaAs/InGaAlAs superlattices [38]. They used a non-equilibrium Green's function approach which is adequate to include both electron wave properties and scattering. They conclude that planar barriers can increase the thermoelectric power factor by a factor of 2.2, but that lateral momentum non-conservation does not improve device performance. This is a little counter-intuitive, if we consider the analogy with optics, where surface microstructuring is used to reduce total internal reflection and increase the amount of light transmission. Bian et al. were able to reproduce the optical results in electron transport using Monte Carlo simulations and elastic scattering at sawtooth interfaces [34]. It is possible that the exact form of momentum scattering plays a role in Wang and Mingo's simulations. Moreover, their power factor calculations for very thick wells (400–500 nm) do not converge to bulk values. In this case, only the elastic scattering process is included. Further simulations and a comparison with experimental cross-plane electrical conductivity and Seebeck coefficient are needed to clarify the role of surface roughness in multilayer structures.

9.4.7 Electron Group Velocity and the Electronic Density of States

Earlier, we discussed the inherent trade-off between electrical conductivity and the Seebeck coefficient in solids. There is also a fundamental trade-off between electronic density of states and electron group velocity in crystalline solids [19]. This is manifested by the fact that solids with a high electron effective mass and/or

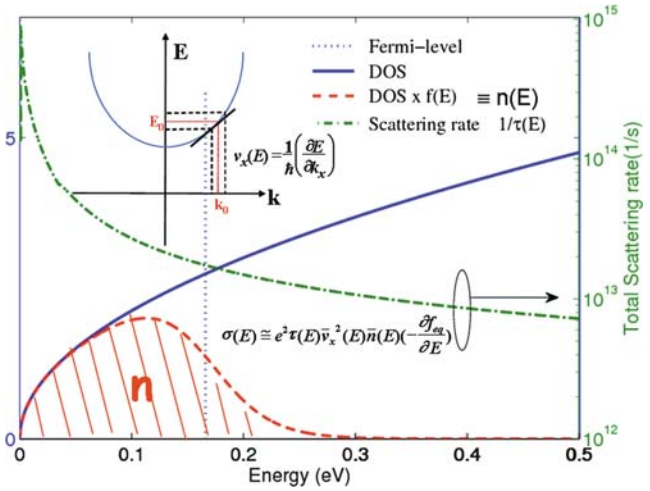

Fig. 9.11 Density of states, scattering rate, and number of electrons in the conduction band vs. energy. Parabolic band structure (E vs. k) is assumed and plotted on the *top left side* of the figure

multivalleys have large densities of states, but at the same time lower mobilities. In Fig. 9.11, we can see that the electronic group velocity is related to the derivative of the dispersion relation (electron energy versus momentum), while the density of states is related to the inverse of the band curvature [7]. The shape of the density of states dominates the overall performance of thermoelectric and thermionic devices, and materials with heavy electron masses and multiple valleys have large 'material' figures of merit and good potential for high ZT [106, 115]. Low-dimensional thermoelectrics and solid-state thermionics try to increase the asymmetry of the differential conductivity by modifying the density of states and the electron scattering, respectively. However, one should remember that the electron group velocity can also be modified, and it is important that the overall product in the differential conductivity should be optimized, rather than each term individually (see Fig. 9.11). Since the density of states is related to the whole dispersion relation in the momentum space, while the electron group velocity is related to the curvature of the band in a given direction, it seems that there should be good opportunities to optimize an ideal *anisotropic* thermoelectric material. Electron effective mass should be small in the direction of transport, while there are lots of states (heavy mass) in the transverse direction. A fundamental study of the trade-off between the sharp increase in the density of states and the large electrical conductivity using hot electron filters is very much needed. Researchers such as N. Mingo and S. Datta have started developing a non-equilibrium Green function formalism for thermoelectric effects in nanostructures [37, 116] (see also Sect. 9.3.4). This allows a first-principles calculation of all the transport coefficients, without assuming any effective medium or other parameters. This could be a basis for designing novel thermoelectric/solid-state thermionic materials from atomic building blocks.

9.4.8 Reversible Thermoelectrics

Recent theoretical analyses by Humphrey and Linke have shown that the electronic efficiency for thermoelectric cooling or power generation can approach the Carnot limit if electron transport between the hot and the cold reservoirs happens in a single energy level under a finite temperature gradient and a finite external voltage [117]. This is in the absence of phonon thermal conductivity and heat losses. Despite ideal conditions, this was an important study, since it showed for the first time what we need to do in order to approach Carnot efficiency in a thermoelectric material (this corresponds to $ZT \rightarrow \infty$). Humphrey and Linke showed that transport at a single optimized energy level will give the maximum current and energy flow, as well as a reduction in electronic thermal conductivity. The latter corresponds to breakdown of the Wiedermann–Franz law, which relates thermal and electrical conductivities of electrons. This could be achieved with an appropriately designed embedded quantum dot material having a graded composition or dot sizes from the hot to the cold junctions. The basic idea is that, whenever there is a finite energy band in which electrical conduction happens, we could have counter-propagating electrical currents from the hot to the cold and from the cold to the hot reservoirs. These currents do not contribute to the net electrical conduction, but they can transport energy from the hot to the cold reservoir (i.e., electronic thermal conductivity). When the electron transport in the material happens at a single energy level, its value can be chosen so that the probability of occupation is identical in both hot and cold reservoirs (see Fig. 9.12). The reservoirs at different temperatures and electrochemical potentials are then in 'energy specific' equilibrium through the material, and there is no net current. Under such conditions, and neglecting the lattice contribution to thermal conductivity, it is possible to achieve thermoelectric energy generation or refrigeration with an efficiency approaching the Carnot limit.

9.5 Reduction of Phonon Thermal Conductivity

Although phonons do not contribute directly to the energy conversion, the reduction of their contribution to the thermal conductivity is a central issue in thermoelectrics research. Several significant increases in the ZT of bulk materials were due to the introduction of thermal conductivity reduction strategies, such as the alloying [9] and phonon rattler concepts [118]. Size effects on phonon transport have long been established, since the pioneering work by Casimir [119] at low temperatures. Since the 1980s, thermal conductivity reduction in thin films has been drawing increasing attention.

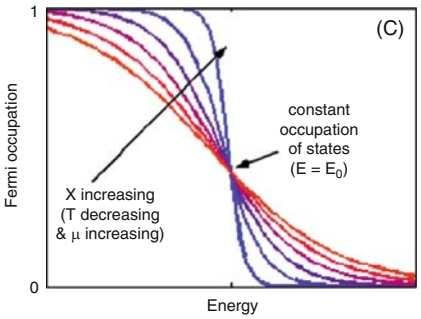

Fig. 9.12 Fermi occupation versus electron energy at various locations between the hot and the cold junctions at an optimum open circuit voltage. There is a specific energy E_0 at which the occupation probability is the same at the hot and the cold junctions anywhere in between. If all the electron transport happens at this specific energy, then one can approach the Carnot limit in thermoelectric energy conversion, provided that the lattice thermal conductivity is negligible. (Courtesy of Dr. Tammy Humphrey and Professor Heiner Linke)

9.5.1 Thermal Conductivity of Superlattices

One proposed approach is to use the thermal conductivity in the direction perpendicular to the superlattice film plane, or the cross-plane direction, while maintaining a low electronic band-edge offset, or ideally, no offset at all [18]. This would allow electron transport across the interfaces without much scattering, while phonons would be scattered at the interfaces [120]. Some early experimental data [121, 122] indicate that the thermal conductivity of superlattices could be significantly reduced, especially in the cross-plane direction. Tien and Chen [123] have suggested the possibility of making superthermal insulators out of superlattices. Extensive experimental data on the thermal conductivity of various superlatttices have been reported in recent years [120–134], mostly in the cross-plane direction. Following such a strategy, Venkatasubramanian's group has reported Bi_2Te_3/Se_2Te_3 superlattices claiming $ZT \sim 2.4$ at room temperature [18].

The mechanisms responsible for reducing thermal conductivity in low dimensional structures have thus become a topic of considerable debate over the last few years. There have been many studies of the phonon spectrum and transport in superlattices since the original work by Narayanamurti et al. [135], but these studies focused on the phonon modes rather than on heat conduction. The first theoretical model predicted a small reduction of the superlattice thermal conductivity [136], due to the formation of minigaps or stop bands. This predicted reduction, however, was too small compared to experimental results in recent years. Two major theoretical approaches were developed in the 1990s to explain the experimental results. One is based on solving the Boltzmann equation with the interfaces of the superlattice treated as boundary conditions [137–140]. The other is based on a lattice dynamics calculation of the phonon spectrum and the corresponding change in the phonon group

velocity [141–145]. More recently, there have also been attempts to use molecular dynamics to simulate the thermal conductivity of superlattices directly [146, 147].

As for electron transport in superlattice structures, there could be several different regimes of phonon transport: the totally coherent regime, the totally incoherent regime, and the partially coherent regime. Lattice dynamics lies in the totally coherent regime. Such approaches are based on the harmonic force interaction hypothesis and thus do not consider anharmonic effects. A bulk relaxation time is often assumed. The main result from lattice dynamics models is that the phonon group velocity reduction caused by the spectrum change can lower the thermal conductivity by a factor of $\sim$ 7–10 at room temperature for Si/Ge superlattices, and by a factor of 3 for GaAs/AlAs superlattices. Although it can be claimed that the predicted reduction in the Si/Ge system is of the order of magnitude that is observed experimentally, the prediction clearly cannot explain the experimental results for GaAs/AlAs superlattices. The lattice dynamics model also shows that, when the layers are 1–3 atomic layers thick, there is a recovery of the thermal conductivity. The acoustic wave model [148], which treats the superlattice as an inhomogeneous medium, shows a similar trend. It reveals that the thermal conductivity recovery is due to phonon tunneling and that the major source of the computed thermal conductivity reduction in the lattice dynamics model is total internal reflection, which in the phonon spectrum representation causes a group velocity reduction. For the experimental results so far, the explanation of the thermal conductivity reduction based on the group velocity reduction has not been satisfactory, even for the cross-plane direction. For the in-plane direction, the group velocity reduction alone leads to only a small reduction in thermal conductivity [145], and cannot explain the experimental data on GaAs/AlAs and Si/Ge superlattices [121, 124, 134]. There is a possibility that the change in the phonon spectrum creates a change in the relaxation time [40], but such a mechanism is unlikely to explain the experimental results for relatively thick-period superlattices, since the density of states does not change in these structures [145].

Models based on the Boltzmann equation which treat phonons as particles transporting heat in inhomogeneous layers lie in the totally incoherent regime [137, 138, 140]. Theoretical calculations have been able to explain the experimental data quantitatively. The models are based on the solution of the Boltzmann equation using the relaxation time in the bulk materials for each layer. Phonon reflection and transmission at the interfaces are modeled on the basis of past studies of the thermal boundary resistance. Compared to lattice dynamics and acoustic wave models, the particle model allows incorporation of diffuse interface scattering of phonons. In the models presented so far, the contribution of diffuse scattering has been left as a fitting parameter. One argument for the validity of the particle model is that thermal phonons have a short thermal wavelength, which is a measure of the coherence properties of broadband phonons inside the solid [137]. It is more likely, however, that the diffuse interface scattering, if it does indeed happen as the models suggest, destroys the coherence of monochromatic phonons and thus prevents the formation of superlattice phonon modes. The particle-based model can capture the effects of total internal reflection, which is partially responsible for the large group velocity reduction under the lattice dynamics models. Approximate methods

to incorporate phonon confinement or inelastic boundary scattering have also been proposed. From the existing modeling, it can be concluded that, for heat flow parallel to the interfaces, diffuse interface scattering is the key factor causing thermal conductivity reduction. For the case of heat conduction perpendicular to the interfaces, phonon reflection, confinement, and also diffuse scattering can greatly reduce the heat transfer and thermal conductivity. The larger the reflection coefficient, the larger the thermal conductivity reduction in the cross-plane direction.

A key unsolved issue concerns the actual mechanisms of phonon scattering at the interfaces, and in particular, what causes diffuse phonon scattering. Phonon scattering has been studied quite extensively in the past in the context of thermal boundary resistance. Superlattice structures grown by epitaxial techniques usually have better interface morphology than the other types of interfaces studied previously. Even for the best material system such as GaAs/AlAs, however, the interfaces are not perfect. There is interface mixing and there are also regions with monolayer thickness variations. These are naturally considered as potential sources of diffuse interface scattering, and a simple model has been developed by Ziman [150]. Another possibility is the anharmonic force between the atoms in two adjacent layers. Models based on the Boltzmann equation assume a constant parameter p to represent the fraction of phonons specularly scattered. Ju and Goodson [149] used an approximate frequency-dependent expression for p given by Ziman [150] in the interpretation of the thermal conductivity of single-layer silicon films. Chen [140] also argued that inelastic scattering occurring at interfaces can provide a path for the escape of confined phonons. A promising approach to resolve this issue is molecular dynamics simulation [146, 147]. In addition to the interface scattering mechanisms, there are also several other unanswered questions. For example, experimental data obtained by Venkatasubrmanian seems to indicate a butterfly-shaped thermal conductivity curve as a function of thickness [18, 151]. Quantitative modeling of the stress and dislocation effects also needs to be further refined.

Since the lattice dynamics and particle models present the totally coherent and totally incoherent regimes, a theoretical approach that can include both effects should be sought. Simkin and Mahan [152] proposed a new lattice dynamics model by the introduction of an imaginary wave vector that is related to the mean free path. This approach leads to the prediction of a minimum in the thermal conductivity value as a function of the superlattice period thickness. For thicknesses greater than the minimum, the thermal conductivity increases with thickness and eventually approaches the bulk values. For thicknesses thinner than the minimum, the thermal conductivity recovers to a higher value. It should be pointed out, however, that the imaginary wave vector represents an absorption process, not exactly a scattering process, as is clear in the meaning of the extinction coefficient of the optical constants. It will be interesting to see whether such an approach can explain the experimentally observed trends of thermal conductivity reduction along the in-plane direction.

9.5.2 Thermal Conductivity of Nanowires

Aside from superlattices and thin films, other low-dimensional structures such as quantum wires and quantum dots are also being considered for thermoelectric applications. There are a few experimental and theoretical studies on the thermal conductivity of quantum dot arrays and nanostructured porous media [153, 154]. Theoretically, one can expect a larger thermal conductivity reduction in quantum wires compared to thin films [155, 156]. Measurements of the thermal conductivity in quantum wires have been challenging. Recent measurements of the thermal conductivity of carbon nanotubes provide a possible approach for measurements on nanowires for thermoelectric applications [157]. The suspended microheater bridge approach has been used quite extensively by Li Shi et al. to characterize, not only the thermal conductivity, but also the electrical conductivity and the Seebeck coefficient of individual nanowires [158]. Four-point measurements (i.e., having two electrodes at each end of the nanowire) can be used to eliminate electrical and thermal contact resistances. Two recent reports highlight the potential of rough silicon nanowires, where thermal conductivity was reduced by two orders of magnitude with much smaller reduction in the electrical conductivity, resulting in ZT values approaching 1 near room temperature [25, 159]. There have been questions about the accuracy of these single nanowire measurements [160]. From a theoretical point of view, the potential role of phonon localization has been mentioned. Mingo et al. have recently ruled out the possibility of observing phonon localization in some nanowires [161]. Recent simulations by Martin et al. have shown that the correct treatment of phonon boundary scattering, which takes into account phonon frequency dependence, can explain the observed low thermal conductivities in rough nanowires 20–50 nm in diameter [161]. In fact, they predicted that the thermal conductivity should depend on the square of the nanowire diameter over the mean roughness. This dependence was observed in earlier measurements.

9.6 Applications

9.6.1 Heterostructure Integrated Thermoelectric/Thermionic Microrefrigerators on a Chip

Using the idea of heterostructure electron energy filtering, thin film coolers based on various materials have been fabricated and characterized. InGaAsP/InP [100, 101, 163], and InGaAs/InP [164], were grown by metal organic chemical vapor deposition (MOCVD), and InGaAs/InAlAs [165], InGaAsSb/InGaAs [166], SiGe/Si [167, 168], and SiGeC/Si [99]) were grown by molecular beam epitaxy (MBE). These structures were lattice matched to either InP or silicon substrates to ease their monolithic integration. Si-based heterostructures are particularly useful for monolithic integration with silicon-based electronics. The basic idea was to use

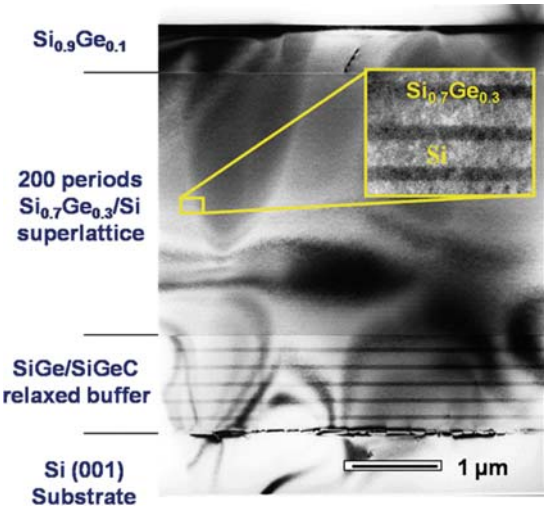

Fig. 9.13 Transmission electron micrograph of 3 μm thick 200× (5 nm $Si_{0.7}Ge_{0.3}$/10 nm Si) superlattice grown symmetrically strained on a buffer layer designed so that the in-plane lattice constant was approximately that of relaxed $Si_{0.9}Ge_{0.1}$. The n-type doping level (Sb) is 2×10^{19} cm^{-3}. The relaxed buffer layer has a ten-layer structure, alternating between 150 nm $Si_{0.9}Ge_{0.1}$ and 50 nm $Si_{0.845}Ge_{0.150}C_{0.005}$. A 0.3 μm $Si_{0.9}Ge_{0.1}$ cap layer was grown with a high doping to get a good Ohmic contact [162]

a band offset between the different layers as a hot carrier filter. The superlattice structure also has the potential to reduce the lattice thermal conductivity. Different superlattice periods (5–30 nm), dopings (1×10^{15}–7×10^{19} cm^{-3}), and thicknesses (1–7 μm) were analyzed. A typical SiGe/Si microrefrigerator shown in Fig. 9.13 consists of a 3 μm thick superlattice layer with a 200× (3 nm Si/12 nm $Si_{0.75}Ge_{0.25}$) structure doped to 5×10^{19} cm^{-3}, a 1 μm $Si_{0.8}Ge_{0.2}$ buffer layer with the same doping concentration as the superlattice, and a 0.3 μm $Si_{0.8}Ge_{0.2}$ cap layer with a doping concentration of 1.9×10^{20} cm^{-3}. Various microrefrigerator devices were fabricated using standard thin film processing technology (photolithography, wet and dry etching, and metallization). In the single-leg microcooler geometry, a gold or aluminum metal contact is used to send current to the cold side of the device (see Fig. 9.14). The Joule heating and heat conduction in this metal layer have a strong impact on the overall cooler performance. An electrical contact on the rear side of the silicon substrate, or on the front surface far away from the device, is used as the second electrode. Thus, three-dimensional heat and current spreading in the substrate helped the localized cooling of the device.

Figure 9.15 shows a scanning electron micrograph of thin film coolers of various sizes (40×40 to 100×100 μm^2). Figure 9.16 illustrates typical cooling curves (maximum cooling below ambient versus supplied current) for 60×60 μm^2 microrefrigerators. For comparison, results are shown for identical devices based on bulk silicon and two different superlattice periods [168, 169]. The bulk Peltier effect in silicon can produce $< 1°C$ cooling, while superlattice structures can increase

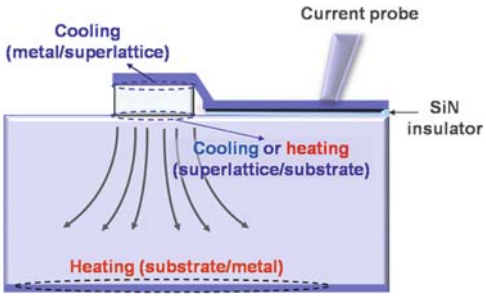

Fig. 9.14 Diagram showing current flow and heat exchange at various junctions in a single-element microrefrigerator on a conductive substrate

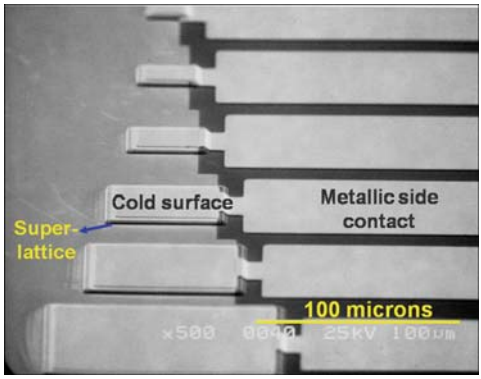

Fig. 9.15 Scanning electron micrograph of thin film coolers of the various sizes in the range from 40×40 to 100×100 μm^2 [162]

the performance to $> 4°C$. Increasing the current, thermoelectric cooling increases linearly, but at some point Joule heating, which is proportional to the square of the current, dominates, and the net cooling is reduced. Figure 9.17 shows the experimental and theoretical cooling for different sizes of microrefrigerator. Calculations are based on commercial finite element 3D electrothermal simulations in which thermoelectric cooling and heating with an effective Seebeck coefficient have been added [170]. Figure 9.18 shows the calculated temperature distribution of a 60×60 μm^2 device at its maximum cooling at room temperature. Figure 9.19 shows the thermal image of a microrefrigerator under operation. One can see uniform cooling on top of the device. No significant temperature rise can be seen on the metal contact layer adjacent to the device. A ring of localized heating around the device is attributed to Joule heating in the buffer layer beneath the superlattice [171].

In Fig. 9.17 we can see that, due to non-ideal effects (Joule heating in the substrate, at the metal–semiconductor junction, and in the top metal contact layer), there is an optimum device size on the order of 50–70 μm in diameter that achieves maximum cooling [172, 173]. This is due to the fact that various parameters scale differently with the device size. For example, both the substrate's 3D thermal and electrical resistances scale as the square root of the device area, while the Joule

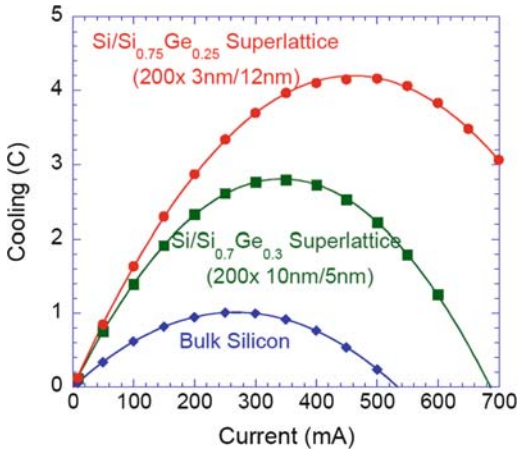

Fig. 9.16 Cooling versus supplied current for bulk silicon microrefrigerator and for 3 μm thick superlattice devices with two different superlattice periods. Device size is 60×60 μm^2 [162]

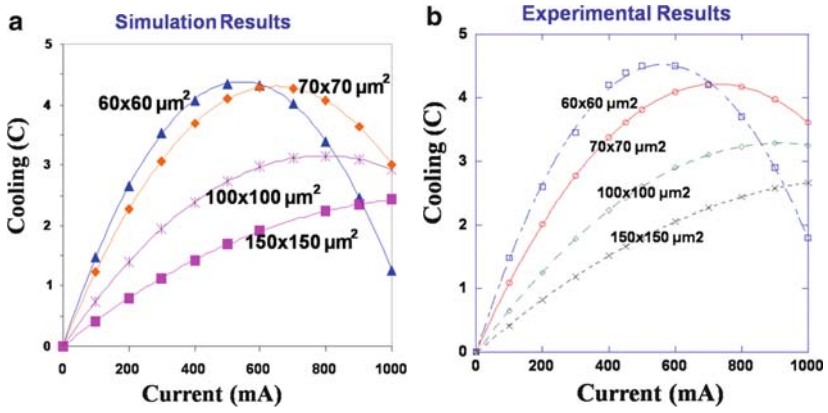

Fig. 9.17 Theoretical and experimental cooling versus supplied current for different microrefrigerator sizes. Microcooler devices consist of a 3 μm thick superlattice layer with the structure of 200× (3 nm Si/12 nm $Si_{0.75}Ge_{0.25}$) and a doping concentration of 5×10^{19} cm^{-3}, a 1 μm $Si_{0.8}Ge_{0.2}$ buffer layer with the same doping concentration as the superlattice, and a 0.3 μm $Si_{0.8}Ge_{0.2}$ cap layer with a doping concentration of 1.9×10^{20} cm^{-3} [162]

heating from the metal–semiconductor contact resistance scales directly proportionally to it [171–173]. The cooling temperature in these miniature refrigerators was measured using two techniques. First, a small $\sim$ 25–50 μm in diameter type E thermocouple is placed on top of the device and another thermocouple is placed farther away, on the heat sink. Even though the thermocouple had the same diameter as the refrigerator, accurate temperature measurements with ~ 0.01°C resolution were achieved on devices with diameter larger than 50–60 μm. We also used an integrated thin film resistor sensor on top of the microcooler. To electrically isolate the

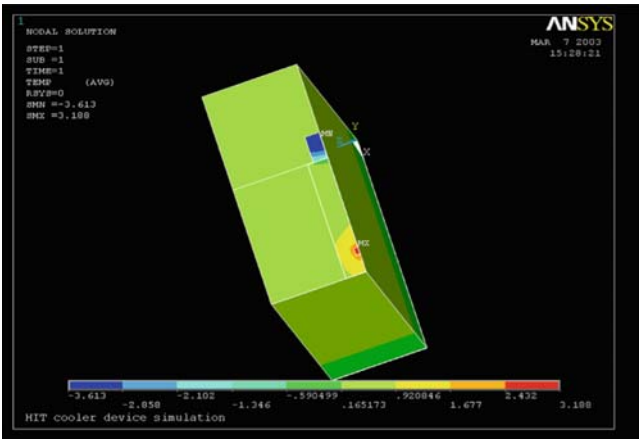

Fig. 9.18 3D electrothermal simulation showing temperature distribution in the SiGe microrefrigerator under bias [162]

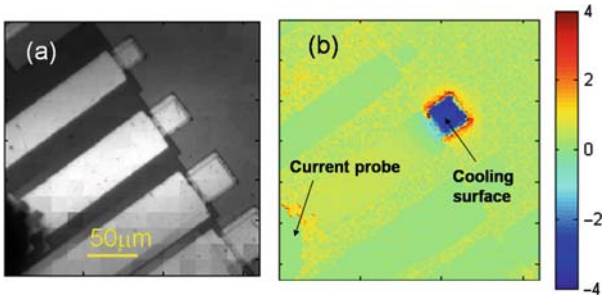

Fig. 9.19 Thermal image of a 50×50 μm^2 microrefrigerator at an applied current of 500 mA. The stage temperature is 30°C and the device is cycled at a frequency of $\sim$ 1 kHz [162]

thin film resistor, a 0.1–0.3 μm thick SiN layer was deposited on the top electrode of the microrefrigerator. The resistance versus temperature was calibrated on a variable temperature heating stage and this was used to measure cooling on top of the device. A resistor could also be used as heat load directly on top of the device if a large current is applied (see Fig. 9.20). The experimental results shown in Fig. 9.20 illustrate the cooling temperature of a 40×40 μm^2 microrefrigerator device as a function of the heat load density. During these measurements, we heat the heater using a constant current, and at the same time we also measure the cooling of the microrefrigerators using a thermocouple or the resistance value of the heater. By increasing the constant current to the heaters, more heat load was added on top of the refrigerators, and the cooling ΔT was decreased. The maximum cooling power density of the device was defined as the maximum heat flux per area that the device could absorb when $\Delta T = 0$. The maximum cooling power density for different microrefrigerators with device sizes 40×40 to 100×100 μm^2 is 600–120 W/cm^2, as indicated in Fig. 9.21.

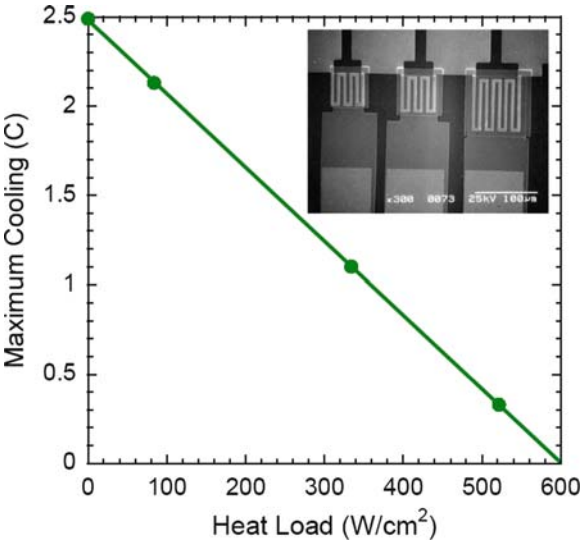

Fig. 9.20 Maximum cooling temperature versus heat load. *Inset* shows the fabricated thin-film heater on top of the SiGe superlattice microrefrigerator [162]

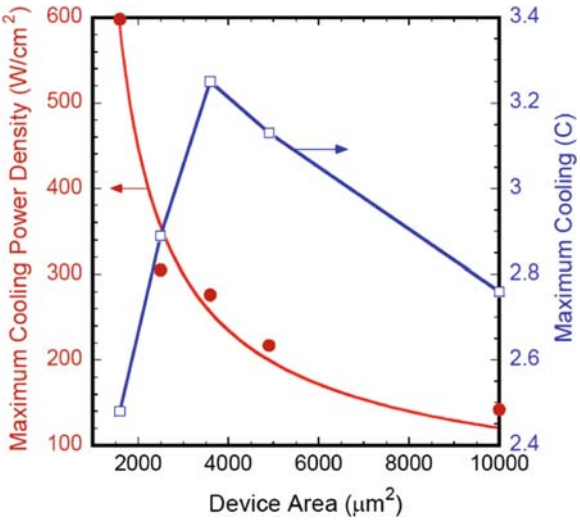

Fig. 9.21 The maximum cooling power density at zero net cooling (*solid circles*) and the maximum cooling temperature at zero heat load (*open squares*) versus device size for the SiGe superlattice microrefrigerator [162]

It is interesting to note that, contrary to the maximum cooling temperature results, the smallest samples ($\sim$ 30–40 μm in diameter) had the largest cooling power densities [174]. This was explained using theoretical models. It is due to the fact that certain parasitic mechanisms, e.g., heat conduction from the heat sink to the cold junction through the metal contact layer, will reduce maximum cooling below

the ambient temperature, while in fact this can improve the cooling power density of the microrefrigerator by creating additional paths for heat spreading.

A metal contact attached directly to the cold surface of the microcooler is a source of parasitic Joule heating and heat leakage to the sink. This limits the maximum cooling of the device by 10–30% [173]. However, these single-leg devices are much simpler to make than conventional thermoelectric coolers, where an array of n-type and p-type semiconductors are used electrically in series and thermally in parallel. The cooling power density is only a function of the element leg length, and it is independent of the number of elements. The main reason for choosing array structures is that they benefit from reduced heat loss in the metal leads, which stems from the trade-off between operating current and voltage. If the goal is to remove a small hot spot, a single-element thermoelectric cooler is much easier to integrate on top of a chip.

Bulk SiGe has a ZT value that is 5–7 times smaller than BiTe at room temperature. III–V semiconductors also have a very low ZT of about 0.01–0.05 [74, 175]. The main use of the HIT coolers mentioned above is not to achieve high efficiencies in order to cool big macroscopic size chips. The key idea is to selectively cool small regions of the chip, removing hot spots locally. If a small fraction of the chip power is dissipated in localized regions, low thermoelectric efficiency is not the most important factor. It is more critical to incorporate small size refrigerators with high cooling power density, and with minimum additional thermal resistances inside the chip package.

When comparing HIT microcoolers with bulk thermoelectric modules, there are three primary advantages. First of all, both micro-size and standard lithographic fabrication methods make HIT refrigerators suitable for monolithic integration inside IC chips. It is possible to put the refrigerator near the device and cool the hot spot directly. The 3D geometry of a device with a small size cold junction and large size hot junction allows heat spreading from the small hot region to the heat sink [172]. Secondly, the high cooling power density surpasses that of commercial bulk TE refrigerators. In fact, the directly measured cooling power density, a figure exceeding 680 W/cm^2 [174], is one of the highest numbers reported so far [60]. Thirdly, the transient response of the current SiGe/Si superlattice refrigerator is much faster than that of bulk TE refrigerators. The standard commercial TE refrigerator has a response on the order of a few tens of seconds. The measured transient response of a typical HIT superlattice sample is on the order of $\sim$ 20–40 μs, again very similar to that of BiTe/PbTe superlattices [176, 177]. Thus microcoolers could be used to remove dynamic hot spots in the chip.

According to the theoretical simulation, the current limitation of superlattice coolers comes from the resistance of the buffer or metal/semiconductor contact layer, which is on the order of 10^{-6} Ωcm^2 [172, 173, 178]. Mingo et al. have recently suggested that it is possible to increase the ZT of a SiGe alloy by embedding silicide nanoparticles with optimum size in the 5–10 nm range [38]. This calculation predicts a room temperature $ZT \sim 0.5$, which can enable monolithic cooling of devices by 15–20°C and cooling power densities exceeding 1000 W/cm^2.

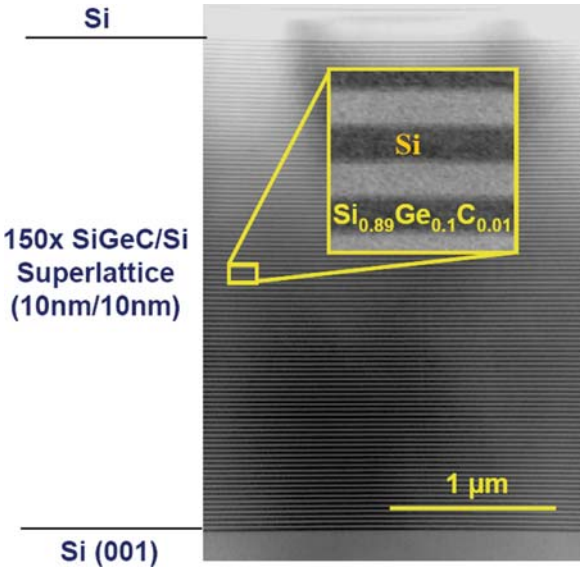

Fig. 9.22 Transmission electron micrograph of an $Si_{0.89}Ge_{0.1}C_{0.01}$/Si superlattice structure grown directly on a silicon substrate [162]

9.6.2 SiGe and SiGeC Superlattice Optimization

Microrefrigerators have been demonstrated based on superlattices of $Si_{1-x}Ge_x$/Si [167, 168], $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ [182], and Si/Ge [179]. Since the lattice constant of $Si_{1-x}Ge_x$ ($x > 0.1$) is substantially different from that of the silicon substrate, a graded buffer layer was used in order to gradually change the lattice constant to that of the average value of the two superlattice layers. This buffer layer, which is $\sim$ 1–2 μm thick, can accommodate lots of dislocations, and it allows growth of a very high quality 3–5 μm thick superlattice on top of it. Maximum cooling of 4.5°C at room temperature, 7°C at 100°C [168, 179], and 14°C at 250°C have been demonstrated [180]. Detailed thermal imaging of these structures has shown that Joule heating in the buffer layer is one of the key non-ideal effects that limit the maximum performance [171, 173, 181]. In addition, since the average lattice constant of the superlattice corresponds to an SiGe alloy with $x \sim 0.1$–0.2, only electronic devices based on SiGe could be monolithically integrated on top of these refrigerators.

The addition of 1–2% carbon to SiGe can decrease its lattice constant and make it match that of silicon. High quality 3 μm thick SiGeC/Si superlattices have been grown without any buffer layer (see Fig. 9.22). Room temperature cooling was lower, about 2.5°C for a 60×60 μm^2 device [99]. However, the cooling performance increased substantially with temperature, and 7°C cooling at 100°C ambient temperature was similar to the best SiGe/Si superlattice devices (see Fig. 9.23).

An important question concerns the role of the superlattice and its effect on the thermoelectric figure of merit ZT. Superlattice structures could lower lattice

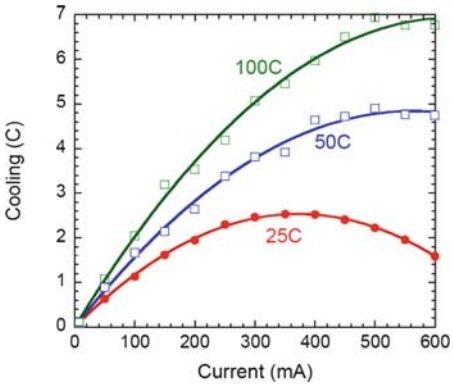

Fig. 9.23 Cooling versus current at different ambient temperatures for $Si_{0.89}Ge_{0.1}C_{0.01}$/Si superlattice microrefrigerators [162]

thermal conductivity and increase the thermoelectric power factor (Seebeck coefficient squared times electrical conductivity). Huxtable et al. characterized the thermal conductivity of various SiGe superlattices [182]. The 3ω technique was used to measure the thin film thermal conductivity in the direction perpendicular to the superlattice plane [183]. The thermal conductivity scaled almost linearly with the interface density, and approached that of the alloy SiGe, but was never lower than that of the alloy ($\sim$ 8–9 W/mK). With a larger difference in the germanium content of the layers, e.g., with $Si_{0.2}Ge_{0.8}/Si_{0.8}Ge_{0.2}$ and Si/Ge, a larger acoustic impedance mismatch could be achieved. This resulted in a lower lattice thermal conductivity than the alloy ($\sim$ 3 W/mK) [182]. However, there were large amounts of dislocations in the 3 μm thick sample, and these reduced the electrical conductivity. Microrefrigerator devices based on this material with low thermal conductivity did not show substantial cooling (only $\sim 1^{\circ}$C), so a good crystalline quality of the SiGe material is essential for high thermoelectric performance [179].

Full microrefrigerator devices based on a bulk thin-film SiGe alloy and based on an SiGe/Si superlattice were fabricated and their cooling characterized. Similar metallization and device geometry were used in order to facilitate the comparison between material properties. Room temperature cooling of the superlattice was about 5% larger than for the bulk alloy film [184]. Given the fact that the thermal conductivity of the alloy was 25% lower than the superlattice (measured independently using the 3ω technique), we estimate that the hot electron filtering in the superlattice increased the thermoelectric power factor by $\sim$ 30% [185]. This shows that, unless techniques are found for suppressing the lattice thermal conductivity of SiGe superlattices below that of alloys (without degradation in the electrical performance), the SiGe alloy films have good cooling performance compared to superlattices, and they may be easier to fabricate and integrate on top of silicon chips [186]. An interesting new direction is the potential to use embedded silicide nanoparticles in an SiGe alloy, which could reduce the room temperature lattice thermal conductivity to $\sim$ 1–2 W/mK without degrading the electrical conductivity [38].

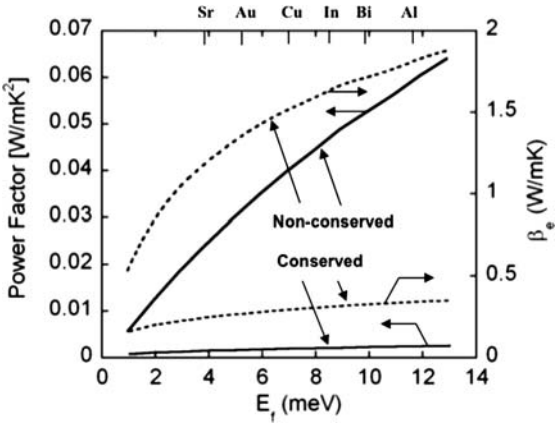

Fig. 9.24 Thermoelectric power factor (*left-hand axis*) and electronic contribution to thermal conductivity (*right-hand axis*) vs. Fermi energy in electronvolts for a metallic superlattice. Optimum barrier height is assumed

9.6.3 Potential Metal/Semiconductor Heterostructure Systems

The discussion in Sect. 9.4.5 showed the potential of thermionic emission in metallic structures. The introduction of tall barriers inside metals will allow the filtering of hot electrons, whence the Seebeck coefficient and the thermoelectric power factor may be significantly increased. Figure 9.24 shows the calculated thermoelectric power factor ($S^2\sigma$) versus Fermi energy E_f. The electronic contribution to thermal conductivity (κ_e) for maximum power factor is also shown on the right-hand axis. The results are shown for both conserved and non-conserved lateral momentum. In the case of non-conserved lateral momentum, a power factor as high as 0.064 W/mK^2 is predicted (corresponding to a ZT value of 6.7, when the lattice thermal conductivity is 1 W/mK). This is due to the higher electrical conductivity and a Seebeck coefficient that resulted from the asymmetric distribution of transported electrons compared to the Fermi energy. The optimum barrier height for the maximum power factor is given in [109]. The mobility is taken to be 12 cm^2/V s, the value for a typical metal. The thermal conductivity in metals is dominated by the electron thermal conductivity, which is approximately $2.44 \times 10^{-8}\sigma T$ in units of W/mK, according to the Wiedemann–Franz law. However, the electrical conductivity (σ) in a metallic superlattice is low compared to that in the bulk metal, whence the electron thermal conductivity can be comparable to that of phonons in the barrier, as can be seen in Fig. 9.24. Many metals can be grown epitaxially on top of semiconductors. However, growth of high quality semiconductors on top of metals is difficult. There are not many candidate systems for high quality, high electron mobility, metal/semiconductor composites. Work at the Thermionic Energy Conversion Center concentrated on two material systems: the first concerns rare-earth-based III–V semiconductors (such as ErAs:InGaAlAs), and the second, the nitride-based metal/semiconductor multilayers (such as TiN/GaN and ZrWN/ScN [187]).

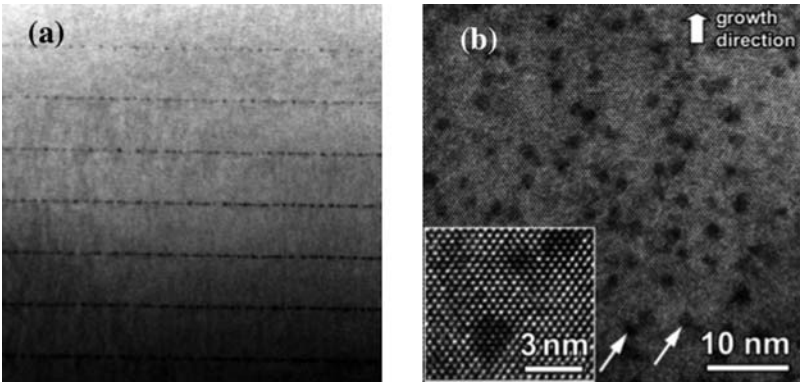

Fig. 9.25 (**a**) Tranmission electron micrograph of ErAs/InGaAs superlattices. In this picture the average ErAs concentration is reduced from 0.4 monolayers to 0.05 monolayer from the bottom to the top of the graph. (**b**) InGaAs matrix with randomly distributed ErAs nanoparticles

9.6.4 InGaAlAs Embedded with ErAs Nanoparticles

ErAs is a rocksalt semimetal which can form into epitaxial nanometer-sized particles on a III–V semiconductor surface. Overgrowth is nucleated on the exposed semiconductor surface between the particles and is essentially defect-free. The properties of the resulting nanocomposite depend on the composition of the host semiconductor and on the particle morphology, which can be controlled during growth. For thermoelectric applications, we concentrated on the incorporation of ErAs into various compositions of InGaAlAs (lattice-matched to InP). The particles pin the Fermi level at an energy that is dependent on both the particle size and the composition of the semiconductor. For example, the Fermi level of InGaAs is pinned within the conduction band, increasing the free electron concentration, and thus the electrical conductivity. This means that ErAs nanoparticles contribute electrons to the conduction band of the host matrix, and make the material n-type. We first focused on developing structures which consisted of superlattices of ErAs islands in an InGaAs matrix, which was lattice-matched to an InP substrate. To maintain a constant ErAs concentration, our initial samples consisted of ErAs depositions ranging from 0.05 monolayers/period to 0.4 monolayers/period, with the superlattice period varying from 5 to 40 nm. While InGaAs is not a good thermoelectric material to start with (room temperature $ZT \sim 0.05$), the incorporation of ErAs reduced the thermal conductivity of the material by approximately a factor of 2 (i.e., total thermal conductivity ~ 4 W/mK) [188]. At the same time, in-plane measurements of the Hall effect showed an increased carrier concentration for smaller particles and a high-quality material with mobilities of 2000–4000 cm^2/V s at 300 K [114].

We then concentrated on the growth of codeposited (randomly distributed) ErAs:InGaAs, which has the advantage of growing much faster than superlattice structures, because it does not require growth interrupts (see Fig. 9.25). This allows us to grow much thicker structures with greater stability. Our initial efforts focused

on 0.3% ErAs, which is the same concentration as the first superlattice samples. The Hall effect and Seebeck measurements have shown that these materials are electrically very similar to the superlattice materials with a thermoelectric power factor similar to bulk InGaAs [114]. On the other hand, the thermal conductivity was reduced by 25% compared to the ErAs/InGaAs superlattice material [188]. The significant reduction compared to the bulk alloy material is due to the increased scattering of mid- to long-wavelength phonons by embedded nanoparticles. This makes this system one of the few materials in which thermal conductivity is reduced below the so-called alloy limit without creating defects that lower electron mobility and electrical conductivity. Kim et al. have recently developed a detailed model for phonon transport in these structures, and the simulated lattice thermal conductivity matches well with the experimental result over a wide temperature range [189]. The measured power factor of the material at room temperature was slightly increased, which resulted in the value of the ZT more than doubling (see Fig. 9.26).

In order to increase the number of carriers participating in transport and improve the thermoelectric power factor, we studied n-type ErAs:InGaAs structures with increased doping and InGaAlAs barriers for electron filtering. These barriers actually consist of a short-period superlattice or 'digital alloy' of InGaAs and InAlAs. By carefully choosing the composition of the InGaAlAs/InGaAs multilayers, we can create electron-filtering barriers to improve the thermoelectric power factor at a given temperature. The cross-plane thermoelectric transport properties were measured using mesa structures with integrated thin film heaters/sensors. Experimental results confirmed the increase in the cross-plane Seebeck coefficient by a factor of three compared to the in-plane value [191, 192].

Recently, we have focused on the incorporation of ErAs into InGaAlAs alloys. The main idea was that the Fermi level pinning at the interface of ErAs/InGaAlAs can be used to create 3D Schottky potential barriers which can selectively scatter hot electrons. This can create a solid-state thermionic device without the use of superlattice barriers. Zebarjadi et al. [193] have developed a Boltzmann-based theory to simulate electron transport in such structures. They included scatterings from phonons, impurities, binary electrons, and the alloy deformation potential. Then nanoparticle scattering rates are added to the other rates. The nanoparticles are investigated in different regimes [194]. When nanoparticle sizes are small and their potential is weak, the Born approximation can be used. This approximation is based on perturbation theory. Zebarjadi et al. showed that the results of the Born approach are valid only for high-energy electrons with energies several times higher than the potential strength. The scattering cross-section of single particles can be calculated exactly by solving the Schrödinger equation inside and outside of the nanoparticle and matching the slope of the wave function at the boundary of the nanoparticle and the host matrix. This method is called the partial-wave method. If one then averages over the fluctuations of the potential (size or strength fluctuations), then the method is called the average T-matrix method. This is valid for low volume fractions of nanoparticles (less than 0.5%). When the nanoparticles are close to each other one needs to include the effect of multiple scatterings. One way to include multiple scatterings is through the effective medium theory. Nanoparticles form a random

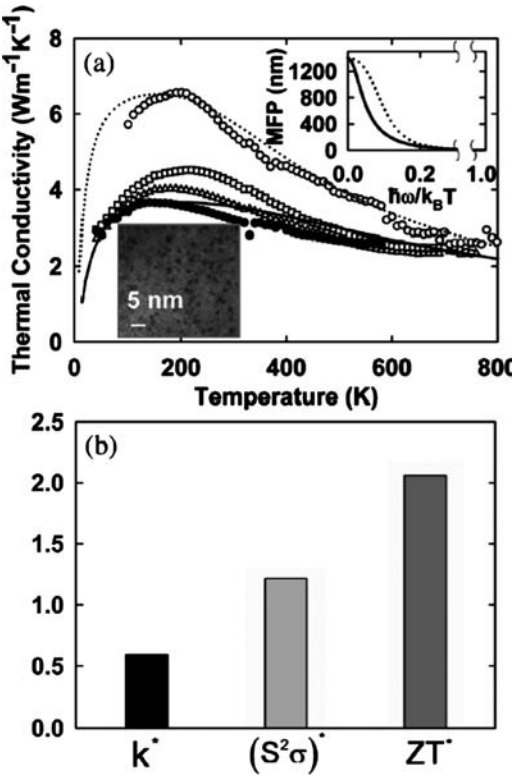

Fig. 9.26 (**a**) Thermal conductivity of randomly distributed ErAs in $In_{0.53}Ga_{0.47}As$ (*solid circles*). Thermal conductivity of $In_{0.53}Ga_{0.47}As$ alloy (*open circles*), 0.4 monolayer with a 40 nm period thickness ErAs:$In_{0.53}Ga_{0.47}As$ superlattice (*open squares*), and 0.1 monolayer with a 10 nm period thickness ErAs:$In_{0.53}Ga_{0.47}As$ superlattice (*open upward triangles*) are shown as references. *Dotted* and *solid lines* are based on theoretical analysis. One *inset* shows TEM pictures of randomly distributed ErAs in $In_{0.53}Ga_{0.47}As$. The other *inset* shows the phonon mean free path (MFP) versus normalized frequency at 300 K. (**b**) Resulting enhancement of the thermoelectric figure of merit at 300 K. Thermal conductivity, power factor, and figure of merit *ZT* of randomly distributed ErAs in $In_{0.53}Ga_{0.47}As$ are normalized by the corresponding values of $In_{0.53}Ga_{0.47}As$ [190]

medium. Electrons move in this medium with their energy plus a self-energy. So on the average the bottom of the conduction band moves with the amount of the self-energy and the band structure is modified. This method is called the coherent potential approximation (CPA). Results of the CPA converge to those of the average T-matrix when the volume fraction of the nanoparticles is small. For higher fractions (15%), the difference can be up to 100 percent. Figure 9.27 shows the effect of different scatterings on the mobility of the Er-doped InGaAlAs sample.

Figure 9.28 shows a comparison between the experimental data and the theoretical predictions for the electrical conductivity and the Seebeck coefficient. The results of the modeling suggest that it will be very challenging to increase the power

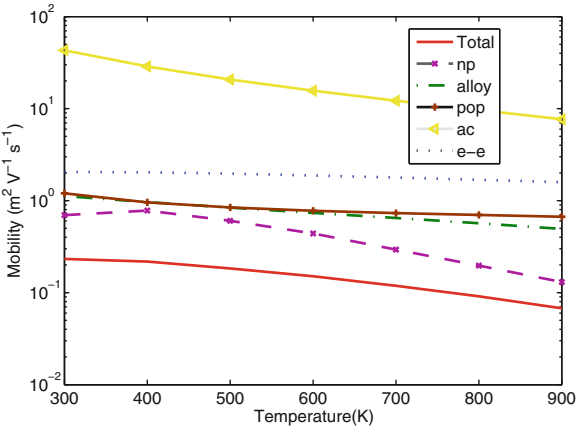

Fig. 9.27 Mobility calculation for individual scattering mechanisms in the $ErAs_{0.003}$:InGaAlAs sample. Aluminum concentration is 20%. Nanoparticle scattering has been calculated using the partial-wave method, and dominates over the whole temperature range

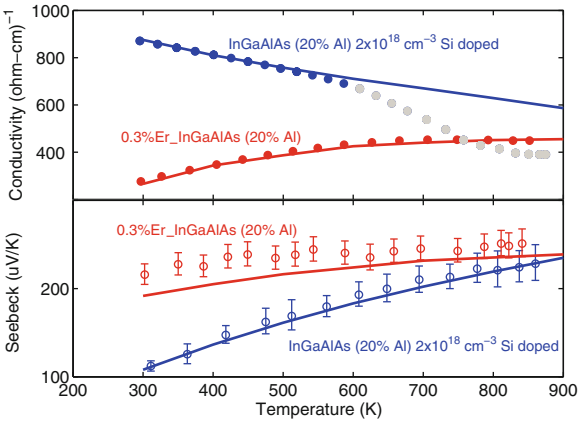

Fig. 9.28 Comparison of theory with experiment. *Circles* are experimental data and *solid lines* are theoretical predictions. Electrical conductivities were fitted (using the alloy potential for the doped sample, and using a_0 and the Schottky barrier height for the nanoparticle sample as fitting parameters), and the Seebeck coefficients were predicted. Experimental electrical conductivity data above 600 K for the sample without nanoparticles were not reproducible (*gray circles*). This is probably due to insufficient passivation and sample degradation at high temperatures

factor significantly. Careful design of the nanoparticle potential and adjustment of the Fermi level are required. One important parameter is the average potential strength, which is the strength of the individual nanoparticles multiplied by their volume fraction. As the volume fraction of nanoparticles increases, more carriers are required to obtain an optimized power factor. This is possible, for example, by co-doping the samples with Si.

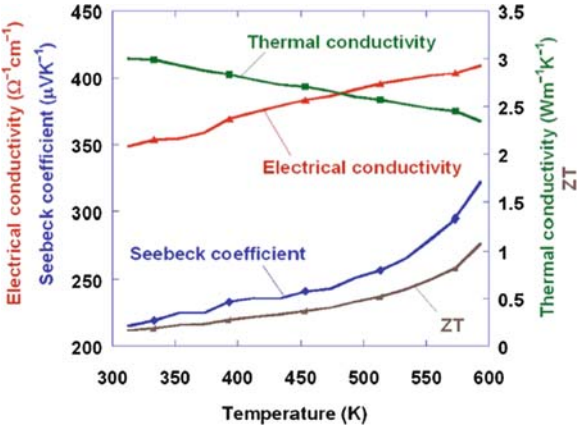

Fig. 9.29 The measured thermal conductivity, electrical conductivity, Seebeck coefficient, and *ZT* of 0.3% ErAs nanoparticles inside an InGaAlAs matrix (20% Al concentration) [195]

As the ErAs:InGaAlAs material is isotropic, measurement of transport properties is much easier. The differential 3ω method was used to measure the thermal conductivity of InGaAlAs (20% Al) embedded with 0.3% ErAs nanoparticles. Figure 9.29 shows the measured thermal conductivity versus temperature. The thermal conductivity decreases with temperature, and the fitting curve is very close to a straight line in the temperature range between 300 and 600 K. The thermal conductivity of ErAs–InGaAlAs (20% Al) is much lower than that of bulk InGaAlAs and very close to that of ErAs:InGaAs.

The electrical conductivity of 0.5 μm thick ErAs:InGaAlAs (20% Al) grown on an insulating InP substrate was measured using the Van der Pauw method (see Fig. 9.29). The electrical conductivity increases with temperature. This is because the number of free electrons thermally excited out of ErAs particles increases with temperature by almost a factor of 3. This was verified by the Hall measurements by Thierry Caillat at JPL.

It is very interesting to see that all three parameters, viz., thermal conductivity, electrical conductivity, and Seebeck coefficient, go in the favored direction of having a larger *ZT* when the temperature increases. This is not usual for bulk materials, and it is due to the ErAs nanoparticles and their hetero-interfaces with the InGaAlAs alloy. The thermoelectric power factor and figure of merit *ZT* were calculated from the three independently measured parameters and plotted in Fig. 9.29. $ZT \sim 1$ is achieved at 600 K. Further measurements are underway to study *ZT* at higher temperatures. The material seems to be stable at temperatures as high as 800 K. However, electrical measurements are affected by the electrical conductivity of the intrinsic InP substrate at higher temperatures. Substrate removal is needed in order to obtain reliable results.

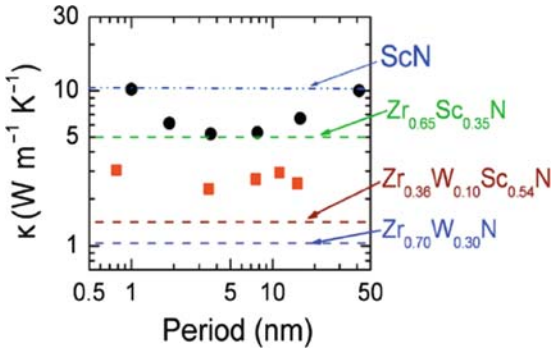

Fig. 9.30 Cross-plane thermal conductivity of 300 nm thick $Zr_{0.64}W_{0.36}N$/ScN (*squares*) and ZrN/ScN (*dots*) multilayers. Superimposed on the plot are *horizontal lines* corresponding to the experimentally determined lattice component of thermal conductivity (i.e., the alloy limit) of different alloys of ZrN, ScN, and W_2N, namely, $Zr_{0.65}Sc_{0.35}N$, $Zr_{0.36}W_{0.10}Sc_{0.54}N$, and $Zr_{0.70}W_{0.30}N$. The data points have an *error bar* that is equivalent to the size of the markers used to represent the measurement result [196]

9.6.5 Metal/Semiconductor Multilayers Based on Nitrides

As an alternative approach to rare-earth nanocomposites, the Thermionic Energy Conversion team decided to explore the rocksalt-structured nitrides, a class of materials that had not been previously investigated for metal/semiconductor epitaxy. The key advantage is the possibility of making full metal/semiconductor multilayers that offer greater control in implementing hot electron filtering. Moreover, the material should be stable at very high temperatures. The rocksalt nitrides include several semiconducting phases, including ScN and a high-pressure polymorph of GaN. There are also several metallic transition metal nitrides that have the conductivity of good metals (15–50 $\mu\Omega$ cm), including TiN and ZrN. As a class, these materials also offer exceptionally high thermal and chemical stability, with melting points typically above 2500°C, and a high degree of oxidation resistance at elevated temperatures. Much of the early work on nitrides focused on pseudomorphic rocksalt GaN stabilized in superlattices with TiN and VN. Although such structures were successfully demonstrated for the first time, the effective critical thickness for rocksalt stabilization (relative to the transformation to the wurtzite phase) was found to be 1–2 nm, too small to prevent excessive tunneling through the semiconductor barriers. Recently, GaN has been substituted with ScN, a semiconducting nitride phase that adopts the rocksalt structure at atmospheric pressure. Combined with a lattice-matched metallic (Zr,W)N alloy, these metal/semiconductor superlattices can be grown with any period from 1 nm and higher by reactive sputter deposition from elemental metallic targets at substrate temperatures of approximately 850°C.

The room temperature thermal conductivity of ScN/(Zr,W)N superlattices has recently been assessed using the time domain photothermal reflectance technique in collaboration with Yee Kan Koh and Professor David Cahill at UIUC (see Fig. 9.30) [196]. A clear minimum in thermal conductivity is revealed at a period of $\sim$ 3–7 nm

for ScN/ZrN. Note that the minimum value of the total cross-plane thermal conductivity (~ 5 W/m K) is well below the thermal conductivity of the constituent materials (the total measured thermal conductivity of ZrN is 47 W/m K with a calculated lattice contribution of 18.7 W/m K). By alloying with W-N to decrease the lattice mismatch with ScN, the thermal conductivity is further reduced to ~ 2.2 W/m K at a period of 6 nm. At higher temperatures ($> 300^{\circ}$C), Umklapp scattering is expected to dominate, and lattice thermal conductivities below 2 W/m K are expected. The cross-plane Seebeck coefficient, power factor, and transient ZT measurements are in progress. Preliminary results give a conduction band offset of 0.96 eV and a Fermi energy of 0.69 eV for ScN (6 nm)/ZrN (4 nm) superlattices [197]. A room temperature Seebeck coefficient of 840 μV/K has been measured, combining the transient I–V measurement and thermal imaging. This system can have ZT values higher than 2 at temperatures above 1000 K, if lateral momentum is not conserved.

9.7 Scaling up Production

Novel metallic-based superlattices with embedded nanoparticles are synthesized by molecular beam epitaxy (MBE), metal organic vapor phase epitaxy (MOCVD), or pulsed laser deposition systems. These techniques allow a precise layer-by-layer growth with a growth rate of 0.1–2 μm per hour. Large-scale MBE growth of GaAs chips for cell phones and laser diodes for compact disc applications have been demonstrated [198]. The epitaxial growth is done simultaneously on 5–6 wafers, each 2–4 inches in diameter. Once the research phase is completed and electronic and thermal properties of nanostructured materials optimized, other techniques such as chemical vapor deposition (CVD) could also be used for larger scale production of nanoengineered multilayer or embedded nanostructure thermoelectric materials. In CVD, growth rates in excess of 100 μm/hr can be achieved. In the case of embedded nanoparticles, once the optimum composition and size have been identified, one may even consider bulk growth techniques, such as the Bridgeman technique. Nanoparticles could be formed under the right thermodynamic conditions to yield the balance between surface energy and mixing free energy.

The Boston College/MIT team has achieved excellent performance in nanopowdered Bi_2Te_3 and SiGe materials, with ZT values reaching 1.5 and 1, respectively [12]. Most of the improvement is achieved by reducing the lattice thermal conductivity without affecting the thermoelectric power factor. Interesting changes in optimum doping and in the peak power factor versus temperature are observed. These are attributed to hot electron filtering at grain boundaries. This team has built a single thermoelectric couple using the nanopowdered material in one leg, and has achieved room temperature cooling of $\sim 100^{\circ}$C.

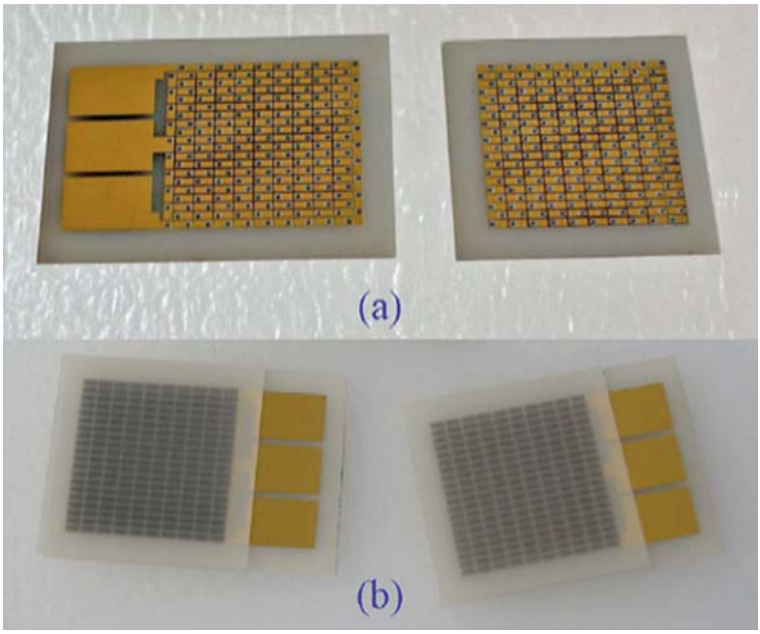

Fig. 9.31 (**a**) 200 n-type elements of ErAs:$(InGaAs)_{0.8}(InAlAs)_{0.2}$ and 200 p-type elements of ErAs:InGaAs were bonded on a lower and upper AlN plate. (**b**) Two 400 element thermoelectric generator modules [195]

9.7.1 Thin-Film Power Generation Modules

To generate a large enough open-circuit voltage, many n-type and p-type thermoelectric elements need to be connected electrically in series and thermally in parallel. We used InGaAlAs alloys embedded with ErAs nanoparticles. The n-type had 20% aluminum and was not intentionally doped. All of the free electrons came from ErAs nanoparticles. The p-type leg had 0% aluminum concentration (i.e., it was InGaAs), and it was doped with Be to reach a free hole concentration of 5×10^{19} cm^{-3}. Wafer scale processing and flip-chip bonding were used to fabricate a multi-element thin-film power generation module. Both n- and p-type InGaAlAs thin films grown on InP substrates were patterned to a 200 element array. Each element had a cross-sectional area of 120×120 μm^2. The element mesas were formed using inductive coupling plasma dry etching. The n- and p-element arrays were flip-chip bonded to the gold-plated AlN substrate. After removing the InP substrate by selective wet etching, the two AlN plates were flip-chip bonded together to form a power generation module. The detailed fabrication process can be found in [199].

Modules with 10 or 20 μm thick thin films were made (see Fig. 9.31). It is also very important to optimize the heat sink so that a large temperature drop can be obtained across the active legs. In this measurement, the heat sink was made of copper with forced cooling water. The generator module was placed on the heat sink,

and heat was applied to the top surface through a copper bar. Thermocouples were used to measure the temperature drop across the generator module. The open-circuit voltages were 2.1 and 3.5 V for modules with 10 and 20 μm tall elements, respectively. The corresponding external temperature drop was 120 K. Variable external load resistances were then used to extract the output power. The maximum output powers per unit area of the active element were 1 and 2.5 W/cm^2, respectively (see Fig. 9.31).

9.7.2 Optoelectronic and Electronic Applications

Thermoelectric microdevices have some immediate applications. If the reported *ZT* is further confirmed and enhanced, the applications will undoubtly expand into many areas. Here we discuss a number of potential applications:

1. temperature stabilization,
2. high cooling density spot cooling, and
3. micropower generation.

Temperature stabilization is very important for optoelectronic devices such as laser sources, switching/routing elements, and detectors requiring careful control over their operating temperature. This is especially true in current high speed and wavelength division multiplexed (WDM) optical communication networks. Long haul optical transmission systems operating around 1.55 μm typically use erbium-doped fiber amplifiers (EDFAs), and are restricted in the wavelengths they can use due to the finite bandwidth of these amplifiers. As more channels are packed into this wavelength window, the spacing between adjacent channels becomes smaller, and wavelength drift becomes very important. Temperature variations are the primary cause of wavelength drift, and they also affect the threshold current and output power in laser sources. Most stable sources such as distributed feedback (DFB) lasers and vertical cavity surface emitting lasers (VCSELs) can generate large heat power densities on the order of kW/cm^2 over areas as small as 100 μm^2 [200, 201]. The output power for a typical DFB laser changes by approximately 0.4 dB/°C. Typical temperature-dependent wavelength shifts for these laser sources are on the order of 0.1 nm/°C [202]. Therefore a temperature change of only a few degrees in a WDM system with a channel spacing of 0.2–0.4 nm would be enough to switch data from one channel to the adjacent one, and even less of a temperature change could dramatically increase the crosstalk between two channels. Temperature stabilization or refrigeration is commonly performed with conventional thermoelectric (TE) coolers. However, since their integration with optoelectronic devices is difficult [200, 203], component cost is greatly increased because of packaging. The reliability and lifetime of packaged modules are also usually limited by their TE coolers [204]. Microdevices monolithically integrated with the functioning optoelectronic devices have advantages over separate devices in terms of their response time, size, and costs.

Many electronic and optoelectronic devices dissipate high heat fluxes. Conventional thermoelectric devices cannot handle large heat fluxes. With reduced leg length, the cooling heat flux of thermoelectric devices increases, thus providing the opportunity to handle high heat flux devices. It should be remembered, however, that more heat flux must be rejected at the hot side and removed using conventional heat transfer technologies, such as heat pipes and high thermal conductivity heat spreaders. The active cooling method is beneficial only when the device needs to be operated below ambient temperatures or for temperature stabilization. Examples are infrared detectors and quantum cascade lasers. The speed of many electronic devices increases with reduced temperature, whence it is possible to use thermoelectric coolers to gain speed. Instead of cooling the whole chip, thermoelectric microcoolers can potentially be applied to handle local hot spots in semiconductor chips [205]. Regions with sizes ranging from a few tens to hundreds of microns in diameter have a temperature 10–30°C higher than the average chip temperature. This causes clock delays and failures in digital circuits. In addition, chip reliability due to electromigration is a thermally activated process, so the mean free time between failures decreases exponentially as the temperature rises.

The use of the Peltier coolers for the thermal management of computer chips has been very limited. The heat dissipation density in IC chips is much larger than the cooling power density of conventional Peltier coolers. Several companies have commercialized thin-film thermoelectric coolers with leg lengths in the 20–200 μm range [206–208]. The highest room temperature cooling power density is ~ 100 W/cm^2, which is close to the average value in IC chips. However, because of the low efficiency of the Peltier device and the power constraints in computer systems, it is still prohibitive to cool the whole chip. Recently, Prasher, Venkatasubramanian, et al. have demonstrated localized cooling of a small millimeter-scale hot spot using thin film Peltier coolers inside the conventional heat sink [209]. The hot spot temperature was reduced by ~ 7°C without affecting the background temperature in the chip. This opens up interesting opportunities for site-specific thermal management in integrated circuits.

Thermoelectric devices have traditionally been used as radiation detectors such as thermopiles, and can be used as power sources. With the rapid developments in MEMS, microscale power supplies have been in increasing demand. Thermoelectric microgenerators can be coupled with environmental heat sources to drive sensors and microdevices for autonomous operation of these devices. The body temperature powered wristwatch is a recent example [210]. Leonov et al. have extensively studied the potential of thermoelectric power generation using body heat [211]. It was shown that $\sim$ 10–100 mW/cm^2 could be extracted. The best location for providing high power was identified to be the head.

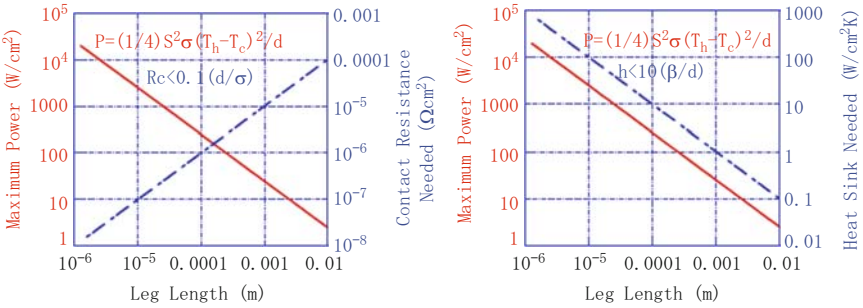

Fig. 9.32 The calculated generated power versus device length d. The figure also shows the minimum contact resistance R_c and the minimum heat transfer coefficient h_c needed so that parasitics do not dominate generator performance

9.8 System Requirements for Power Generation

In order to demonstrate large scale direct thermal-to-electric energy conversion with efficiencies higher than 20%, an average ZT of the material > 1.5 is necessary through the temperature range 300–950 K. This can be achieved by grading the material (e.g., changing the superlattice period, barrier height, doping, nanoparticle size, or composition, etc.) and optimizing the properties to maximize the performance at each local temperature. Power generation density is inversely proportional to the thermoelectric leg length, and a goal of 1 W/cm^2 will require legs shorter than 1 cm. With growth techniques such as molecular beam epitaxy and MOCVD, it is extremely hard to grow thick layers. Thick layers also require a lot of nanostructured material and this is quite expensive. If thinner material is used, higher power densities can be achieved. However, parasitic loss mechanisms could start to dominate. The key factors are electrical contact resistance between the electrodes and the thermoelectric material, and the finite thermal resistance of the heat sink.

Assuming generic material parameters, i.e., Seebeck coefficient $S = 200$ mV/K, electrical conductivity $\sigma = 1000/\Omega\,\text{cm}$, thermal conductivity $\beta = 1$ W/mK, a hot side temperature of $T_h = 900$ K and cold side temperature of $T_c = 400$ K, Fig. 9.32 shows the generated power versus device length d. This figure also shows the minimum contact resistance R_c and the minimum heat transfer coefficient h_c needed so that parasitics do not dominate generator performance (i.e., their contribution is 10% or less). One can see that with 10 μm thick devices, a contact resistance less than $10^{-7}\ \Omega\,\text{cm}^2$ is needed. This is quite possible. On the other hand, a 10 μm device will need a heat sink with a heat transfer coefficient of 100 W/cm^{2}K, otherwise the performance will be significantly degraded. This requirement is several orders of magnitude higher than the best sink demonstrated to date, so heat sinking is quite an important limiting factor. On the other hand, the goal is not to generate more than the 1000 W/cm^2 that an ideal 10 μm thick device could achieve! One can reduce the heat sink requirement by only covering a fraction F of the hot and cold surfaces

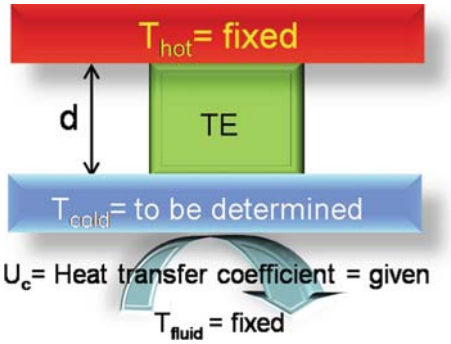

Fig. 9.33 Thermoelectric power generator in which only a fraction of the hot and cold surfaces are covered by the TE element. The cold surface is in contact with a non-ideal heat sink

with the thermoelectric material (see Fig. 9.33). This will limit the flow of the heat to the hot side. Of course, fractional coverage only works when there is good heat spreading at the hot and cold surfaces, so F cannot be too low. With 1% coverage, one can produce several tens of W/cm^2 with a heat sink requirement of 1 W/cm^{2}K.

Assuming that the hot and cold side temperatures of the thermoelectric leg are constant, it is instructive to note that the expression for the conventional power generation density depends only on the thermoelectric power factor, and it increases as the thickness is reduced:

$$P = \frac{1}{4} S^2 \sigma \frac{(T_\mathrm{h} - T_\mathrm{c})^2}{d} \,. \tag{9.20}$$

On the other hand, if we assume a heat sink with finite thermal resistance, the expression for the power generation density will also depend on the material's thermal conductivity, and there is an optimum thickness that gives the maximum power. In the limit of *small temperature gradients*, i.e., $(T_\mathrm{h} - T_\mathrm{c})/4 \ll T_\mathrm{c}$, the following analytical expressions can be derived:

$$d_\mathrm{optimum} \approx \frac{\kappa \left(1 + \frac{1}{2} Z\overline{T}\right) F}{h_\mathrm{c}} \,, \quad P_\mathrm{max} \approx \frac{1}{16} \frac{h_\mathrm{c}}{\overline{T}} \frac{Z\overline{T}}{1 + \frac{1}{2} Z\overline{T}} (T_\mathrm{hot} - T_\mathrm{fluid})^2 \,, \tag{9.21}$$

where $\overline{T} = (T_\mathrm{h} + T_\mathrm{c})/2$. One can see that the optimum thermoelectric material thickness is inversely proportional to the heat transfer coefficient, and it can be reduced by fractional coverage of the surfaces. The maximum power generation density is directly proportional to the heat transfer coefficient. It is a function of the ZT of the material, but it saturates at high ZT values, and more importantly, the maximum power generation density is independent of F as long as heat spreading thermal resistance can be neglected. The above expressions are not a good approximation under large temperature gradients, and for a more accurate analysis a second degree

equation based on the heat balance in the device should be solved. This makes the solution rather less intuitive. However, the main conclusions regarding an optimum module thickness, the effect of fractional coverage, and the importance of the heat sink remain valid. In a situation where heat transfer from the hot source is also a limiting factor, one can show that the above expressions for optimum thickness and maximum power can be generalized by replacing h_c by $h_c h_h/(h_c + h_h)$, where h_h is the heat transfer coefficient with the hot source. It is interesting to note that, when only a fraction of the cold surface is covered by the thermoelectric material and the metal interconnects, one could use the vacant areas and incorporate thermophotovoltaic (TPV) cells that convert the infrared radiation from the hot surface and generate additional electric power.

9.9 Graded Materials

Different thermoelectric materials perform best in different temperature ranges. In a thermoelectric generator under a large temperature gradient, typically the *local ZT is maximized*. Multiple sections with uniform material composition and doping concentration in each section are usually used. These are called functionally graded thermoelectric materials (FGMs) [212]. Beyond maximization of the local ZT, it is found that compatibility among multiple sections must be taken care of, considering that the electrical current is the same and the heat flux is almost continuous along the legs [213]. In a recent study, Bian et al. have gone beyond the conventional approach and shown that the *uniform efficiency criterion* can yield much better performance than optimization of the local ZT [214, 215]. Detailed analytical and numerical simulations have shown that the maximum cooling performance of conventional Bi_2Te_3 materials can be increased by 27% compared to the state-of-the-art using the novel grading approach [214]. The coefficient of performance (efficiency) near maximum cooling can also be significantly increased.

Figure 9.34 compares the Seebeck coefficient profile and the local ZT distribution for the optimal uniform and inhomogeneous materials, respectively, when they are operated at their maximum cooling conditions. The slight changes in the Seebeck coefficient and the ZT of the uniform material are due to the temperature dependence of the material properties. It is interesting that the optimal profile of the inhomogeneous material has a significantly larger Seebeck coefficient but lower ZT near the hot junction.

The idea of continuously graded materials can be applied to both conventional thermoelectric materials and metal–semiconductor nanocomposites for solid-state thermionics. It is thus important to optimize the whole power generation system, rather than just considering the material's local ZT.

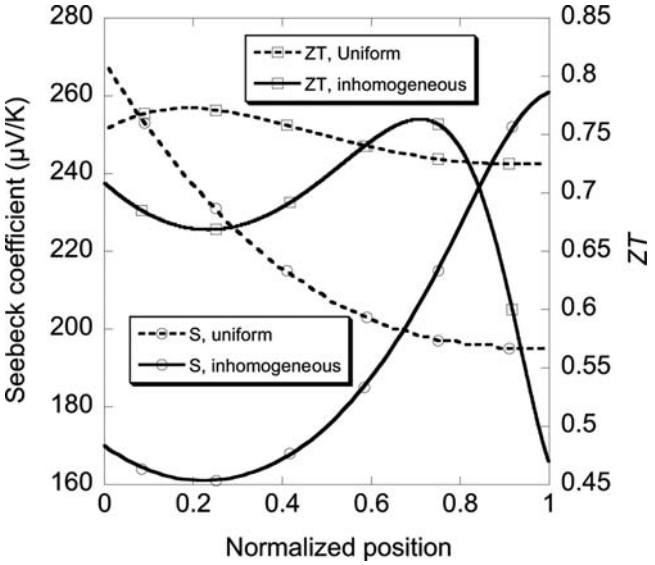

Fig. 9.34 The Seebeck coefficient profile and the local ZT distribution for the optimal uniform and inhomogeneous Bi_2Te_3 materials

9.10 Characterization Techniques

The characterization of thermoelectric properties has turned out to be the most challenging issue for the development of nanostructure-based thermoelectric materials. First, the thermal conductivity measurements are not easy, even for bulk materials, and for thin films, these measurements become considerably more difficult. Even the normally easier measurements in bulk materials, e.g., to obtain the electrical conductivity and Seebeck coefficient, can be complicated due to the small thickness of the film and contributions from the substrate.

It is generally recognized that the thermal conductivity is a difficult parameter to measure. Fortunately, thin-film thermal conductivity measurements have drawn considerable attention over the past two decades, and various methods have been developed. One popular method for measuring the thermal conductivity of thin films is the 3ω method [183, 216]. For thermoelectric thin films such as superlattices, there are several complications. For example, thermoelectric films are semiconductors, so an insulating film is required between the heater and film. The superlattice thermal conductivity is highly anisotropic. The 3ω method is typically applied to measure the cross-plane thermal conductivity by ensuring that the heater width is much greater than the film thickness. There is often an additional buffer layer between the film and the substrate. For Si/Ge, the buffer is graded, and thus has a continuously varying thermal conductivity profile. In applying the 3ω method, there is also the contrast factor that must be considered between the film and the substrate. When the film and the substrate have similar properties, more complicated modeling is needed.

By careful modeling and experimental design, the 3ω method can be applied to a wide range of thin films for measuring the thermal conductivity in both the in-plane and the cross-plane directions [217]. Other methods such as a.c. calorimetry, photothermal methods, and pump-and-probe methods have also been used to measure the thermal diffusivity of superlattices. Detailed reviews of existing methods can be found in [19, 218–221]. By assuming that the specific heat does not change much, which is usually a valid assumption, the thermal conductivity of the structures can be calculated from the measured diffusivity.

In the last couple of years, Cahill et al. have significantly expanded the use of time-domain thermoreflectance (TDTR) [222]. The previous pump–probe transient decay measurements only used the decay in the 1–3 ns range, and tried to fit it using different parameters in the thin-film thermal resistance and the metal transducer/thin film interface boundary resistance. Cahill noticed that, in addition to the femtosecond repetition rate (typically 80 MHz), most setups also include an acousto-optic or electro-optic modulator to chop the signal and take advantage of lock-in detection [223]. The lower frequency modulation (100 kHz–10 MHz) provides information about the thermal penetration in the device at much deeper lengths. Using an ingenious (and somehow mysterious!) ratio of the in-phase and out-of-phase parts of the lock-in signal, Cahill was able to extract the cross-plane thin film thermal conductivity quite accurately for a wide range of materials. By scanning the laser spot, he was also able to provide thermal conductivity maps on the surface of the material [224].

Although measurement of the electrical conductivity and Seebeck coefficient is considered relatively straightforward for bulk materials, it has turned out to be much more complicated for thin films. For transport along the thin film plane, the complications arise from the fact that most thin films are deposited on semiconductor substrates, and the thermoelectric effect of the substrates can overwhelm that of the films. To circumvent these difficulties, several approaches have been taken, such as removing the substrate or growing the film on insulating layers. For example, Si/Ge superlattices are grown on silicon-on-insulator structures. Even with these precautions, there are still complications, such as the existence of the buffer. Thus, differential measurements are sometimes used to subtract the influence of the buffer layer.

9.10.1 Cross-Plane Seebeck Measurement

For transport in the cross-plane direction, measurements of the electrical conductivity and Seebeck coefficient become much more difficult because the films are usually very thin and one cannot use conventional 4-probe or Van der Pauw geometry. We used 50–100 μm diameter mesa structures and integrated thin film heater/sensors on top of the superlattice layer to characterize the cross-plane Seebeck coefficient over a wide range of temperatures (see Fig. 9.35) [165, 225]. The difficulty in characterizing the Seebeck coefficient of a superlattice material lies in

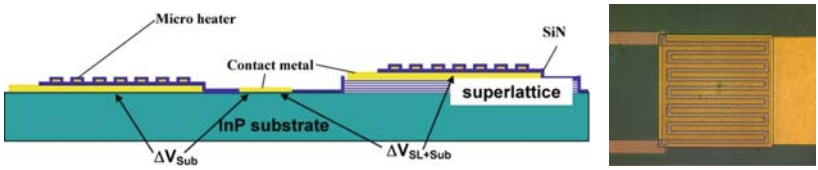

Fig. 9.35 Integrated thin film heater structure used to measure the cross-plane Seebeck coefficient. The electrical contact layer on top of the superlattice and on the substrate allow measurement of the thermoelectric voltage. The thin film resistor acts both as a source of heat and as a temperature sensor. Differential measurements on top of the superlattice and on the substrate are needed in order to extract both the thin film cross-plane Seebeck coefficient and the substrate Seebeck coefficient

simultaneously measuring the voltage and temperature drops to within a few microns on both sides of a thin film. In the above measurement, there could be a significant portion of the temperature drop across the substrate. In order to calculate the substrate contribution, similar thin film heaters were fabricated on a sample where the superlattice was etched away. By using differential measurements, the contribution of the superlattice could be accurately deduced [226]. In addition to the standard DC measurements, where a steady-state temperature gradient is created across the thin film, the 3ω technique has also been used [33]. Similar to the 3ω thermal conductivity measurement, this is a more sensitive technique to estimate the temperature increase across the thin film. However, in addition, the Seebeck voltage generated in the cross-plane direction should also have a 2ω component (proportional to the Joule heating). This allows more accurate measurement of small thermoelectric voltages using lock-in techniques.

9.10.2 Transient ZT Measurement

The recently reported ZT values between 2–3 for Bi_2Te_3/Se_2Te_3 superlattices were obtained using the transient Harman method [18]. Although the method is well established for bulk materials, the application to thin film structures requires careful consideration of various heat losses and heat generation through the leads. In addition, the conventional transient Harman method gives ZT rather than individual thermoelectric properties, such as the Seebeck coefficient.

In order to extract the intrinsic cross-plane ZT of the superlattice by eliminating the effects of the substrate and any parasitics, the bipolar transient Harman technique was used to measure the device ZT of samples with different superlattice thicknesses [60]. High-speed packaging is needed to reduce signal ringing due to any electrical impedance mismatch. Singh et al. achieved a short time resolution of roughly 100 ns in a transient Seebeck voltage measurement [227]. Detailed 3D thermal simulations showed the importance of heat transfer along the leads connected to the top of the superlattice [228]. Due to the large device area compared to the

superlattice thickness, heat transfer along the probes can create a non-uniform temperature distribution on top of the superlattice. Moreover, a significant fraction of the Peltier heating could be transported through the leads. Both of these effects will influence the transient Harman technique and will lead to incorrect ZT values. Once the device and lead geometry had been optimized, accurate ZT measurements could be achieved. The measured intrinsic cross-plane ZT of the ErAs:InGaAs/InGaAlAs superlattice structure with a doping of 1×10^{19} cm^{-3} is 0.13 at room temperature [227]. This value agrees with both calculations based on the Boltzmann transport equation and direct measurements of specific film properties. Theoretical calculations predict that the cross-plane ZT of this superlattice will be greater than 1 at temperatures greater than 700 K.

Recently, the transient Harman technique has been optimized to measure the ZT of the thin film directly [60, 229]. This can work if the parasitic electrical resistance at the metal contact/semiconductor interface is reasonably small. An electrical pulse is applied to the thermoelectric device. This current pulse creates thermoelectric cooling and heating at the junctions and Joule heating in the bulk of the material. Subsequently, a temperature difference develops across the thin film. As the electrical pulse is turned off, the voltage across the device is monitored. The Ohmic voltage drops almost instantaneously (on a sub-ps time scale) while the thermoelectric voltage disappears with the time scale of heat diffusion in the device (10 ns–10 μs for thin films). If the device is under adiabatic conditions (i.e., heat flow from the hot to the cold junction through the contact leads can be neglected), one can obtain the ZT of the device by comparing the thermoelectric voltage pulses when the polarity of the current changes. Joule heating in the material is independent of the current direction, while Peltier cooling or heating at interfaces depends on the current direction [228]. Using this technique, the ZT of BiTe [60] and ErAs:InGaAs/InGaAlAs [227] superlattices have been measured. In addition, the transient Harman technique has been combined with thermoreflectance thermal imaging in order to extract all of the cross-plane thermoelectric properties (σ, S, and κ), as well as the ZT of the thin film [230].

9.10.3 Suspended Heater and Nanowire Characterization

In order to characterize the heat transfer and the thermoelectric transport coefficient of single nanowires, Shi, Majumdar, et al. have developed two suspended thin film heater structures separated by a few microns [231]. A nanowire is placed between the two heaters using an atomic force microscope. In order to make simultaneous electrical and Seebeck coefficient measurements, 2 electrodes are placed on each heater platform. These electrodes allow 4-wire electrical conductivity measurements. The thermal and electrical interface resistances between the nanowire and the metal electrodes on the suspended platform could be large since the nanowire is placed manually. Focused ion beam deposition of metals on the electrodes at the two ends of the nanowire are used to reduce the contact resistances. Once the nanowire

is well attached to the suspended heater platform, thermoelectric measurements can be performed. By sending a current to one of the heaters, a temperature difference of 1–5°C is established across the nanowire. Since both the active and the inactive heaters are thermally insulated from the silicon chip carrier via very long beams, any temperature rise in the inactive heater is due to heat conduction through the nanowire. Hence the thermal conductivity of the nanowire can be extracted. By simultaneously measuring the thermoelectric voltage generated by the nanowire, the Seebeck coefficient can also be extracted. Shi has shown that measurements of thermovoltages across the 4 electrodes (2 on each platform) can be used to estimate the thermal interface resistance between the nanowire and the electrode. This uses the Seebeck coefficient of the nanowire itself as a thermometer. Finally, 4-probe electrical measurement is used to extract the electrical conductivity. Thus all three thermoelectric properties of an individual nanowire can be extracted.

9.11 Thermoelectric/Thermionic vs. Thermophotovoltaics

Thermophotovoltaics (TPV) is a competing technology for direct thermal-to-electric energy conversion. Thermal radiation from a hot source is incident on a filter that transmits only photons at the peak emission [232]. All other photons are reflected back to the hot source. Transmitted photons are converted to electron/hole pairs in a pn-junction diode. Significant losses in conventional photovoltaics [233] are avoided since the diode has a bandgap matching the peak emission of the hot source. TPV cells with efficiencies exceeding 20% have already been demonstrated [234]. They suffer from low power generation densities. Moreover, small bandgap bipolar diodes are very sensitive to non-radiative recombination in the depletion region, Auger recombination, etc. One of the reasons why TPV cells have a higher efficiency than TE or solid-state TI devices is the fact that they have lower parasitic losses. Heat conduction by phonons is a major loss mechanism, since it is electrons that do the work, but in almost all practical thermoelectric materials, the number of free electrons is several orders of magnitude lower than the number of atoms undergoing vibrations and transmitting heat. Metal-based thermionic energy filters have the potential to overcome this problem, and have much higher numbers of free electrons participating in transport.

However, there is another fundamental limit. As pointed out in a lucid paper by Humphrey and Linde [117] (see Sect. 9.4.8), there are inherent electronic thermal conduction losses, since electrons are in contact with both hot and cold reservoirs simultaneously. If the electronic band in the material has a finite width, there is some heat transfer between the two reservoirs, even when there is no net voltage generated. Electrons with energies less than the Fermi energy move from the cold side to the hot side, while electrons with energies higher than the Fermi level move from the hot contact to the cold one. There is no net current, but there is entropy generation [235]. This problem can be overcome if the material is designed in such

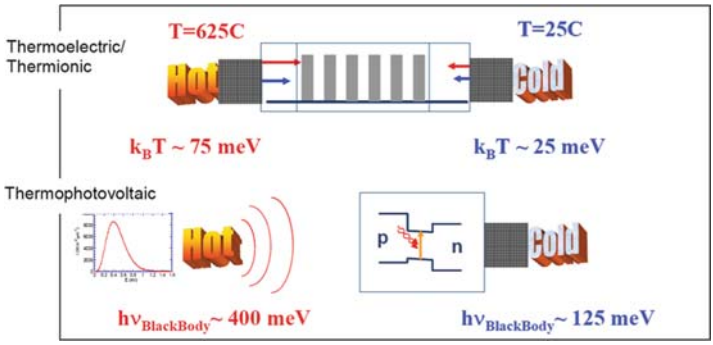

Fig. 9.36 Comparison between TE/TI devices and TPV devices. It is interesting that the average energy of photons exchanged between hot and cold reservoirs is higher than the average energy of electrons exchanged with reservoirs at the same temperature

a way that there is monoenergetic electron transport at a special energy level. This is analogous to the photon filter in TPV devices that transmit only 'good' photons.

Another interesting difference between TE/TI devices and TPV devices is the fact that the average energy of photons exchanged between hot and cold reservoirs is higher than the average energy of electrons exchanged with reservoirs at the same temperature (see Fig. 9.36). The peak in the Planck distribution at, e.g., 900 K, is due to photons with energies of ~ 0.4 eV, while the electron average energy is $\sim 3 \times 0.075 = 0.22$ eV (assuming 3 degrees of freedom). This may seem curious, since the same Carnot limit applies to both electrons and photons. Carnot efficiency is not derived for specific distribution functions, and it is based on general thermodynamic arguments [236]. It seems that working with different energy carriers (electrons, photons, etc.) and with reservoirs with different internal degrees of freedom may provide another opportunity to engineer the efficiency of the heat engines and to approach the entropy limit (second law of thermodynamics) more easily [237, 238].

9.12 Ballistic Electron and Phonon Transport Effects

Electron and phonon transport perpendicular to interfaces raise interesting heat transfer and energy conversion issues. One example is the question of where heat is generated. Joule heating is often treated as uniform volumetric heat generation. In heterostructures, the energy relaxation from electrons to phonons occurs over a distance comparable to the film thickness, and heat generation is no longer uniform. For single-layer devices, this could benefit the device efficiency in principle [239]. Such non-uniform heat generation is a type of hot electron effect that has been studied in electronics [240], and has also been discussed quite extensively in the literature in the context of ultrafast laser–matter interactions [241]. When the size of the thin film

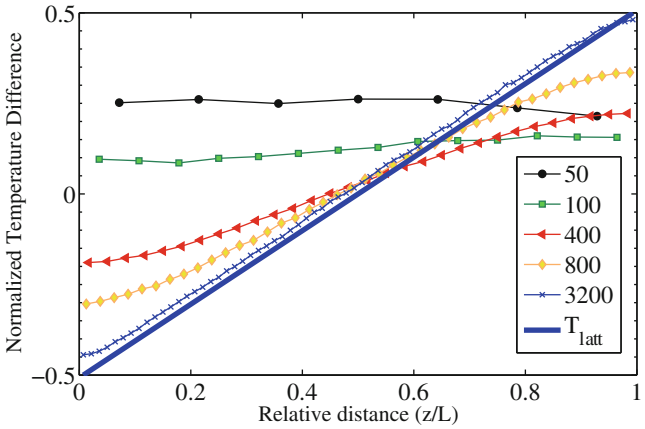

Fig. 9.37 Monte Carlo simulation results. Normalized temperature difference versus relative distance from the left contact. A linear 'lattice' temperature drop is enforced along the sample. This enforced temperature is plotted as a reference. The resulting electronic temperatures are plotted for sample sizes ranging from 50 to 3200 nm

is comparable to the electron energy relaxation length, transport is mainly ballistic, as electrons do not have enough time to relax with the lattice.

Figure 9.37 is the result of a Monte Carlo simulation. In this simulation we enforced a linear lattice temperature drop over a layer of InGaAs. The layer is then placed between two contacts. Electrons were injected from the contact to the InGaAs layer. At small device sizes, the electrons pass the layer ballistically, which results in a flat distribution of electronic temperature. On the other hand, when the layer thickness is large, each electron goes through lots of scatterings and eventually relaxes with the lattice. In the latter case, the electronic temperature tends to the lattice temperature. Another example is the concurrent consideration of ballistic electron transport and ballistic phonon transport, coupled with nonequilibrium electron–phonon interactions.

Zeng and Chen [242] started from the Boltzmann equations for electrons and phonons and obtained approximate solutions for the electron and phonon temperature distributions in heterostructures. In this case, both the electron and the phonon temperatures exhibit a discontinuity at the interface. The phonon temperature discontinuity is the familiar thermal boundary resistance phenomenon. Zeng and Chen concluded that, in the nonlinear transport regime, it is the electron temperature discontinuity at the interface that determines the thermionic effect, and the electron temperature gradient inside the film that determines the thermoelectric effect. On the other hand, calculations by Vashaee and Shakouri [243] assumed a continuous electron temperature across the interface, and focused on the effect of the electron–phonon coupling coefficient in the temperature distribution in HIT coolers. In order to extract the correct boundary condition for the electron temperature at the heterostructure interface, Zebarjadi et al. [35] developed a Monte Carlo code to simulate electron transport in thin film heterostructures. They defined the local quasi-Fermi

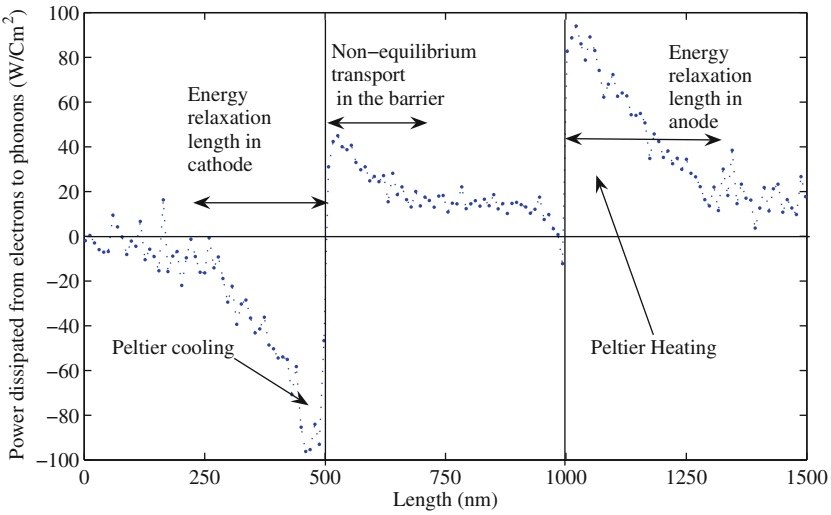

Fig. 9.38 Source term (electron–phonon energy exchange) along the heterostructure, obtained from Monte Carlo simulation. The applied bias is 40 mV. Non-equilibrium regions at the ends of the contacts are disregarded. Peltier cooling can be observed at the cathode–barrier junction, and Peltier heating at the anode–barrier junction

level and electronic temperature from the local population of the electrons. By looking at the electron–phonon exchange energy along a single-barrier structure, they showed that most of the heating happens inside the highly doped contact layers. The Peltier cooling and heating are broadened delta functions inside the contact layers. (see Fig. 9.38)

9.13 Nonlinear Thermoelectric Effects

There are many electronic devices in which charge transport is nonlinear, and one has to go beyond the concept of electrical conductivity [244]. However, nonlinear thermoelectric effects have not been explored to a large extent. The thermoelectric effect at a pn junction is an example of where the bias-dependent Seebeck coefficient can be defined [245]. In the case of nanoscale heat and charge transport in superlattices, quantum wires, and dots, or in point contacts [246], large temperature and electric field gradients and strong interaction of heat and electricity may require one to go to higher order terms in the perturbation of the distribution function [see (9.6)]. This will introduce novel transport coefficients. In this case, even the separation between electrical transport and thermoelectric transport may not be valid, and one has to consider transport coefficients that are a function of both electric field and temperature gradient. A Monte Carlo simulation of the electron distribution function in a device under large currents was used to calculate the bias-dependent

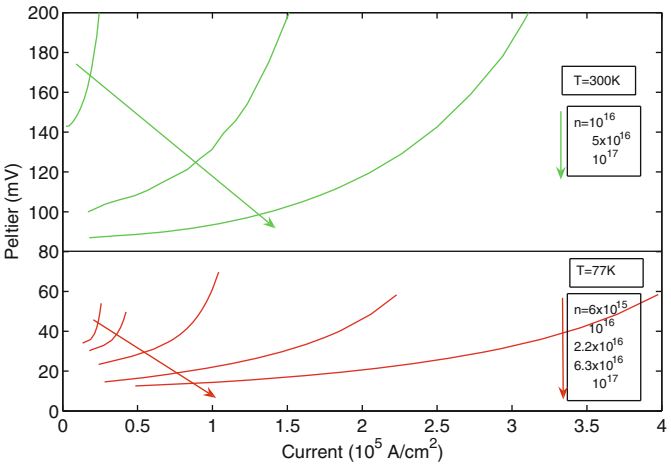

Fig. 9.39 Monte Carlo simulation. Nonlinearity of the Peltier coefficient in n-type InGaAs. Peltier coefficient vs. applied electric current for different doping densities and at two different temperatures. Doping densities are shown on the figure in units of cm^{-3}. The figure shows that nonlinearity is stronger at low doping densities and low temperatures

Peltier coefficient [104] (see Fig. 9.39). Results show that nonlinearity occurs when electronic temperature starts to exceed the lattice temperature. The current threshold at which the Peltier coefficient becomes nonlinear is, e.g., 104 A/cm^2 for InGaAs doped for maximum cooling at 77 K. This current density is achievable experimentally in thin film devices. Detailed calculations show that the nonlinear Peltier effect can improve the cooling performance of thin film microrefrigerators by 700% at 77 K [104].

9.14 A Refrigerator Without the Hot Side

An interesting question is raised by Fig. 9.9, which displays cooling by thermionic emission: is it necessary to have a hot junction at the anode side of the device? By bandgap engineering and appropriate doping, it should be possible to enhance the interaction of electrons with, for example, photons, so that hot carriers at the anode side lose their energy by emitting light rather than heating the lattice. This does not violate the second law of thermodynamics, since the light emission could be incoherent and the total entropy of the electron and photon system would still be increasing. The light emission could occur in a conventional pn junction or in a more elaborate unipolar quantum cascade laser configuration [247]. Calculations by Pipe et al. showed that semiconductor laser structures could be designed to have heterostructure energy filtering near the active region [248]. This can provide internal cooling by several hundred W/cm^2 under typical operating conditions. This method of cooling can be viewed as an electrically-pumped version of the conventional laser cooling which has been used for atom trapping and recently for cooling macroscopic objects [249].

9.15 Conclusion

In this chapter, we have reviewed recent progress in nanostructured thermoelectric and solid-state thermionic energy conversion. Electron energy and current transport can differ significantly from that in bulk materials. The nanoscale size effects and hot electron and phonon filtering can be used to improve the energy conversion efficiency. Recent studies have led to quite a large increase in ZT values and significant new insights into thermoelectric transport in nanostructures. There is, however, much left to be done in new material syntheses, characterization, physical understanding, and new device fabrication. Nonlinear thermoelectric effects and unconventional electron–phonon–photon couplings should provide additional opportunities to make better energy conversion devices.

Acknowledgements The experimental and theoretical data presented in the figures are the results of the work of outstanding students and postdocs: Chris Labounty, Xiaofeng Fan, Gehong Zeng, Daryoosh Vashaee, James Christofferson, Yan Zhang, Zhixi Bian, Kazuhiko Fukutani, Rajeev Singh, Alberto Fitting, Younes Ezzahri, Tela Favaloro, Philip Jackson, Joshua Zide, Je-Hyeong Bahk, Hong Lu, Vijay Rawat, Peter Mayer, Woochul Kim, Suzanne Singer and Scott Huxtable. The authors would like to acknowledge a very fruitful collaboration with Profs. John Bowers, Art Gossard, Susanne Stemmer (UCSB), Arun Majumdar, Peidong Yang (Berkeley), Venky Narayanamurti (Harvard), Rajeev Ram (MIT), Tim Sands (Purdue), Yogi Joshi and Andrei Federov (Georgia Tech), Bob Nemanich (ASU), Avram Bar-Cohen (Maryland), Keivan Esfarjani and Sriram Shastry (UCSC), Stefan Dilhaire (Univ. of Bordeaux), Li Shi (Univ. of Texas), Kevin Pipe (Univ. of Michigan), Joshua Zide (Univ. of Delaware), Ceyhun Bulutay (Bilkent Univ.) Lon Bell (BSST) and Dr. Ed. Croke (HRL Laboratories LLC). This work was supported by DARPA MTO and DSO offices, ONR MURI Thermionic Energy Conversion Center, Packard Foundation, and the Interconnect Focused Center.

References

1. Vining, C.: An inconvenient truth about thermoelectrics, Nature Materials **8**, 83–85 (2009)
2. Van der Sluis, P.: Improvement in cooling stoves, Intl. Patent Pub. No. WO/2006/103613 (2006)
3. Amatya, R., and Ram, R.J.: Materials for Solar Thermoelectric Generators, Materials Research Society Spring Meeting, San Francisco, CA, Talk N1.7 (2009)
4. Goldsmid, H.J.: *Thermoelectric Refrigeration*, Plenum Press, New York (1964)
5. Shastry, S.B.: Electrothermal transport coefficients at finite frequencies, Rep. Prog. Phys. **72**, 016501 (2009)
6. Antokhin, A.Y., and Kozlov, VA.: Relationship between noise characteristics and efficiency of thermoelectric conversion, Soviet Phys. Semiconductors USSR **24**, 926 (1990)
7. Rowe, D.M. (Ed.): *Handbook of Thermoelectrics*, CRC Press, Boca Raton (1995)
8. Humphrey, T.E., O'Dwyer, M.F., and Linke, H.: Power optimization in thermionic devices. J. Phys. D: Appl. Phys. **38**, 2051–2054 (2005)
9. Ioffe, A.F.: *Semiconductor Thermoelements and Thermoelectric Cooling*, Infosearch Limited, London (1957)
10. Tritt, T.M. (Ed.): *Recent Trends in Thermoelectric Materials Research*, In: Semiconductors and Semimetals, Vols. 69–71, Academic Press, San Diego (2001)

11. Slack, G.: In: *CRC Handbook of Thermoelectrics*, ed. by D.M. Rowe, CRC Press, LLC, Boca Raton (1995)
12. Poudel, B., Hao, Q., Ma, Y., Lan, YC., Minnich, A., Yu, B., Yan, X., Wang, D.Z., Muto, A., Vashaee, D., Chen, X.Y., Liu, J.M., Dresselhaus, M.S., Chen, G., and Ren, Z.: High-thermoelectric performance of nanostructured bismuth antimony telluride bulk alloys. Science **320**, 634 (2008)
13. Heremans, J.P. , Jovovic, V., Toberer, E.S., Saramat, A. , Kurosaki, K., Charoenphakdee, A., Yamanaka, S., Snyder, G.J.: Enhancement of thermoelectric efficiency in PbTe by distortion of the electronic density of states, Science **321**, 554 (2008)
14. Hicks, L.D., and Dresselhaus, M.S.: Effect of quantum-well structures on the thermoelectric figure of merit, Physical Review B **47**, 12727–12731 (1993)
15. Dresselhaus, M.S., Lin, Y.M., Cronin, S.B., Rabin, O., Black, M.R., Dresselhaus, G., and Koga, T.: *Quantum Wells and Quantum Wires for Potential Thermoelectric Applications*. In: Semiconductors and Semimetals **71**, 1–121 (2001)
16. Shakouri, A., and Bowers J.E.: Heterostructure integrated thermionic coolers, Applied Physics Letters **71**, 1234–1236 (1997)
17. Mahan, G.D.: *Thermionic Refrigeration*. In: Semiconductor and Semimetals **71**, 157–174 (2001)
18. Venkatasubramanian, R.: *Phonon Blocking Electron Transmitting Superlattice Structures as Advanced Thin Film Thermoelectric Materials*. In: Semiconductors and Semimetals **71**, 175–201 (2001)
19. Chen, G.: *Phonon Transport in Low-Dimensional Structures*. In: Semiconductors and Semimetals **71**, 203–259 (2001)
20. Harman, T.C., Taylor, P.J., Spears, D.L., and Walsh, M.P.: Thermoelectric quantum-dot superlattices with high ZT, J. Electronic Materials **29**, L1–L4 (2000)
21. Dubios, L.: An introduction to the DARPA program in advanced thermoelectric materials and devices, Proc. Int. Conf. Thermoelectrics, ICT'99, 1–4 (1999)
22. Koh, Y.K., Vineis, C.J., Calawa, S.D., Walsh, M.P. and Cahill, D.G.: Lattice thermal conductivity of nanostructured thermoelectric materials based on PbTe, Appl. Phys. Lett. **94**, 153101 (2009); Vineis, C.J., Harman, T.C., Calawa, S.D., Walsh, M.P., Reeder, R.E., Singh, R., and Shakouri, A.: Carrier concentration and temperature dependence of the electronic transport properties of epitaxial PbTe and PbTe/PbSe nanodot, Phys. Rev. B **77**, 235202 (2008)
23. Zhou, F., Szczech, J., Pettes, M.T., Moore, A.L., Jin, S., and Shi L.: Determination of transport properties in chromium disilicide nanowires via combined thermoelectric and structural characterizations, Nano Lett. **7**, 1649 (2007)
24. Fardy, M., Hochbaum, A.I., Goldberger, J., Zhang, M.M. and Yang, P.D.: Synthesis and thermoelectrical characterization of lead chalcogenide nanowires, Adv. Mater. **19**, 3047 (2007)
25. Hochbaum, A.I., Chen, R.K., Delgado, R.D., Liang, W.J., Garnett, E.C, Najarian, M., Majumdar, A., and Yang, P.D.: Enhanced thermoelectric performance of rough silicon nanowires, Nature **451**, 7175 (2008)
26. Shi, L., Yu, C.H., and Zhou, J.H.: Thermal characterization and sensor applications of one-dimensional nanostructures employing microelectromechanical systems, J. Phys. Chem. B **109**, 22102 (2005)
27. Chiritescu, C., Cahill, D.G., Nguyen, N., Johnson, D., Bodapati, A., Keblinski, P., and Zschack, P.: Ultralow thermal conductivity in disordered, layered WSe_2 crystals, Science **315**, 351 (2007)
28. Lundstrom M.: *Fundamentals of Carrier Transport*, Cambridge University Press, Cambridge (2000)
29. Vashaee, D., and Shakouri, A.: Electronic and thermoelectric transport in semiconductor and metallic superlattices, J. of Appl. Phys. **95**, 1233 (2004)
30. Meshkov, S.V.: Tunneling of electrons from a two-dimensional channel into bulk, Sov. Phys. JETP **64** (6), 1337 (1986)
31. Levine, B.F., Bethea, C.G., Hanian, G., Shen, V.O., Pelve, E., Abbott, R.R., and Hsieh, S.J.: Tunable long-wavelength detectors using graded barrier quantum wells grown by electron-beam source molecular-beam epitaxy, Appl. Phys. Lett. **56**, 851 (1990)

32. Vashaee, D., Zhang, Y., Shakouri, A., Zeng, GH., Chiu, Y.J.: Cross-plane Seebeck coefficient in superlattice structures in the miniband conduction regime, Phys. Rev. B **74**, 195315 (2006)
33. Bian, Z., Zebarjadi, M., Singh, R., Ezzahri, Y., Shakouri, A., Zeng, G., Bahk, J.H., Bowers, J.E., Zide, J.M.O., Gossard, A.C.: Cross-plane Seebeck coefficient and Lorenz number in superlattices, Phys. Rev. B **76**, 205311 (2007)
34. Bian, Z.X., and Shakouri, A.: Enhanced solid-state thermionic emission in nonplanar heterostructures, Appl. Phys. Lett. **88**, 012102 (2006)
35. Zebarjadi, M., Shakouri, K., and Esfarjani, K.: Thermoelectric transport perpendicular to thin-film heterostructures calculated using the Monte Carlo technique, Phys. Rev. B **74**, 195331 (2006)
36. Zebarjadi, M., Bulutay, C., Esfarjani, K., and Shakouri, A.: Monte Carlo simulation of electron transport in degenerate and inhomogeneous semiconductors, Appl. Phys. Lett. **90**, 092111 (2007)
37. Paulson, M., and Datta, S.: Thermoelectric effect in molecular electronics, Phys. Rev. B **67**, 241403 (2003)
38. Wang, S., and Mingo, N.: Tailoring interface roughness and superlattice period length in electron-filtering thermoelectric materials, Phys. Rev. B **79**, 115316 (2009); Bulusu and Walker, D.G.: Quantum modeling of thermoelectric properties of Si/Ge/Si superlattices, IEEE Transactions on Electron Devices **55**, 423–429 (2008)
39. Reddy, P., Jang, S.Y., Segalman, R.A., Majumdar, A.: Thermoelectricity in molecular junctions, Science **315**, 1568 (2007)
40. Balandin, A., and Wang, K.L.: Effect of phonon confinement on the thermoelectric figure of merit of quantum wells, J. Applied Physics **84**, 6149–6153 (1998)
41. Walkauskas, S.G., Broido, D.A., Kempa, K., and Reinecke, T.L.: Lattice thermal conductivity of wires, J. Appl. Phys. **85**, 2579 (1999)
42. Chen, Y.F., Li, D.Y., Yang, J.K., Wu, Y.H., Lukes, J.R., and Majumdar, A.: Molecular dynamics study of the lattice thermal conductivity of Kr/Ar superlattice nanowires, Physica B Cond. Matt. **349**, 270 (2004)
43. Klitsner, T., VanCleve, J.E., Fisher, H.E., and Pohl, R.O.: Phonon radiative heat transfer and surface scattering, Phys. Rev. B **38**, 7576–7594 (1988)
44. Chen, Y.F., Li, D.Y., Lukes, J.R., and Majumdar, A.: Monte Carlo simulation of silicon nanowire thermal conductivity, J. Heat Trans. **127**, 1129 (2005)
45. Angelescu, D.E., Cross, M.C., and Roukes, M.L.: Heat transport in mesoscopic systems, Supperlattices and Microstructures **23**, 673 (1998)
46. Heikes, R.R., Ure, R.W.: *Thermoelectricity: Science and Engineering*, Interscience Publishers (1961)
47. Terasaki, I., Sasago, Y., and Uchinokura, K.: Large thermoelectric power in $NaCo_2O_4$ single crystals, Phys. Rev. B **56**, R12685 (1997)
48. Singh, D.J., and Kasinathan, D.: Thermoelectric properties of Na_xCoO_2 and prospects for other oxide thermoelectrics, J. Electronic Mat. **36**, 736 (2007)
49. Koshibae, W., Tsutsui, K., and Maekawa, S.: Thermopower in cobalt oxides, Phys. Rev. B **62**, 6869 (2000)
50. Peterson, M.R., Shastry, B.S., and Haerter, J.O.: Thermoelectric effects in a strongly correlated model for Na_xCoO_2, Phys. Rev. B **76**, 165118 (2007)
51. Ohta H., Sugiura, K., Koumoto K.: Recent progress in oxide thermoelectric materials: p-type $Ca_3Co_4O_9$ and n-type $SrTiO_3$, Inorganic Chemistry **47**, 8429–8436 (2008)
52. Scullin, M.L., Yu, C., Huijben, M., Mukerjee, S., Seidel, J., Zhan, Q., Moore, J., Majumdar, A., and Ramesh, R.: Anomalously large measured thermoelectric power factor in $Sr_{1-x}La_xTiO_3$ thin films due to $SrTiO_3$ substrate reduction, Appl. Phys. Lett. **92**, 202113 (2008)
53. Shakouri, A., Labounty, C., Abraham, P., Piprek, J., and Bowers, J.E.: Enhanced thermionic emission cooling in high barrier superlattice heterostructures, Materials Research Society Symposium Proceedings **545**, 449–458 (1998)

54. Vashaee, D., Shakouri, A.: Conservation of lateral momentum in heterostructure integrated thermionic coolers. Thermoelectric Materials – Research and Applications. Symposium (Materials Research Society Symposium Proceedings Vol. 691). Mater. Res. Soc. pp. 131–145, Warrendale, PA, USA (2001)
55. Shakouri, A.: Nanoscale devices for solid state refrigeration and power generation. Twentieth Annual IEEE Semiconductor Thermal Measurement and Management Symposium, IEEE pp. 1–9, Piscataway, NJ, USA (2004)
56. Shakouri, A.: Thermoelectric, thermionic and thermophotovoltaic energy conversion, International Conference on Thermoelectrics, Clemson, NC, pp. 492–497 (2005)
57. Sun, X., Cronin, S.B., Liu, J.L., Wang, K.L., Koga, T., Dresselhaus, M.S., and Chen, G.: Experimental study of the effect of the quantum well structures on the thermoelectric figure of merit in Si/Si_xGe_{1-x} system, Proceedings of Int. Conf. Thermoelectrics, ICT'99, pp. 652–655 (1999)
58. Hicks, L.D., Harman, T.C., Sun, X., Dresselhaus, M.S.: Experimental study of the effect of quantum-well structures on the thermoelectric figure of merit, Physical Review B **53** (16), R10493–R10496 (1996)
59. Kim, R., Datta, S., and Lundstrom, M.S.: Influence of dimensionality on thermoelectric device performance, J. of Appl. Phys. **105**, 034506 (2009)
60. Venkatasubramanian, R., Siivola, E., Colpitts, T., and O'Quinn, B.: Thin-film thermoelectric devices with high room-temperature figures of merit, Nature **413**, 597–602 (2001)
61. Harman, T.C., Taylor, P.J., Walsh, M.P., and LaForge, B.E.: Quantum dot superlattice thermoelectric materials and devices, Science **297**, 2229–2232 (2002)
62. Hsu, K.F., Loo, S., Guo, F., Chen, F., Dyck, J.S., Uher, C., Hogan, T., Polychroniadis, E.K., Kanatzidis, M.G.: Cubic $AgPb_mSbTe_{2+m}$: Bulk thermoelectric materials with high figure of merit, Science **303** (5659), 818–821 (2004)
63. Broido, D.A. and Reinecke, T.L.: Effect of superlattice structure on the thermoelectric figure of merit, Physical Review B **51**, 13797–13800 (1995)
64. Sofo, J.O., and Mahan, G.D.: Thermoelectric figure of merit of superlattices, Applied Physics Letters **65**, 2690–2692 (1994)
65. Chen, G., and Shakouri, A.: Heat transfer in nanostructures for solid-state energy conversion, Journal of Heat Transfer, Transactions of the ASME **124** (2), 242–252 (2002)
66. Broido, D.A., and Reinecke, T.L.: Theory of thermoelectric power factor in quantum well and quantum wire superlattices, Phys. Rev. B **64**, 045324 (2001)
67. Dresselhaus, M.S., Lin, Y.M., Dresselhaus, G., Sun, X., Zhang, Z., Cronin, S.B., Koga, T., Ying, J.Y.: Advances in 1D and 2D thermoelectric materials. Eighteenth International Conference on Thermoelectrics. Proceedings, (Cat. No.99TH8407). IEEE, pp. 92–99 (1999); Lin, Y.M., and Dresselhaus, M.S.: Thermoelectric properties of superlattice nanowires, Phys. Rev. B **68**, 075304 (2003)
68. Zou, J., and Balandin, A.: Phonon heat conduction in a semiconductor nanowire, Journal of Applied Physics **89** (5), 2932–2938 (2001)
69. Heremans, J.P., Thrush, C.M., Morelli, D.T., Wu, M.-C.: Resistance, magnetoresistance, and thermopower of zinc nanowire composites, Physical Review Letters **91** (7), 076804/1–4, 15 (2003)
70. Deyu, L., Prieto, A.L., Yiying, W., Martin-Gonzalez, M.S., Stacy, A., Sands, T., Gronsky, R., Peidong, Y., Majumdar, A.: Measurements of Bi_2Te_3 nanowire thermal conductivity and Seebeck coefficient, 21st International Conference on Thermoelectrics Proceedings (Cat. No.02TH8657), IEEE, pp. 333–336 (2002)
71. Sander, M.S., Prieto, A.L., Gronsky, R., Sands, T., Stacy, A.M.: Control and assessment of structure and composition in bismuth telluride nanowire arrays. Synthesis, Functional Properties and Applications of Nanostructures. Symposium, Materials Research Society Symposium Proceedings **676**, pp. Y8.35.1–5 (2002)
72. Zhang, Z.B., Dresselhaus, M.S., Ying, J.Y.: Fabrication, characterization and electronic properties of bismuth nanowire systems. Thermoelectric Materials – Next Generation Materials for Small-Scale Refrigeration and Power Generation Applications. Symposium. Mater. Res. Soc. pp. 351–356 (1999)

73. Rabin, O., Herz, P.R., Lin, Y.M., Cronin, S.B., Akinwande, A.I., Dresselhaus, M.S.: Arrays of nanowires on silicon wafers. 21st International Conference on Thermoelectrics Proceedings (Cat. No.02TH8657), IEEE, pp. 276–279 (2002)
74. Mingo, N.: Thermoelectric figure of merit and maximum power factor in III–V semiconductor nanowires, Appl. Phys. Lett. **84**, 2652 (2004)
75. Sharp, J., Thompson, A., Trahey, L., Stacy, A.: Measurement of thermoelectric nanowire array properties. 24th International Conference on Thermoelectrics Proceedings (Cat. No. 05TH8854C) IEEE., pp. 83–86 (2005)
76. Mahan, G.D., and Sofo, J.O.: The best thermoelectric, Proc. National Academy of Sciences of the United States of America **93**, 7436–7439 (1996)
77. Cai, J.W., and Mahan, G.D.: Transport properties of quantum dot arrays, Phys. Rev. B **78**, 035115
78. Hatsopoulos, G.N., and Kaye, J.: Measured thermal efficiencies of a diode configuration of a thermo electron engine, J. Appl. Phys. **29**, 1124–1125 (1958)
79. Mahan G.D.: Thermionic refrigeration, J. Appl. Phys. **76**, 4362 (1994)
80. Mahan, G.D., Woods, L.M.: Multilayer thermionic refrigeration, Physical Review Letters **80** (18), 4016–4019 (1998)
81. Hishinuma, Y., Moyzhes, B.Y., Geballe, T.H., Kenny, T.W.: Vacuum thermionic refrigeration with a semiconductor heterojunction structure, Applied Physics Letters **81**, No. 22, 4242–4244 (2002)
82. Hishinuma Y., Geballe, T.H., Moyzhes, B.Y., and Kenny, T.W.: Measurements of cooling by room-temperature thermionic emission across a nanometer gap, J. Appl. Phys. **94**, 4690 (2003)
83. Gerstenmaier, Y.C., and Wachutka, G.: Efficiency of thermionic and thermoelectric converters. Thermophotovoltaic Generation of Electricity: TPV7, AIP Conference Proceeding **890**, 349 (2007)
84. Tsu, R., Greene, R.F.: Inverse Nottingham effect cooling in semiconductors. Electrochemical & Solid-State Letters **2** (12), 645–647 (1999)
85. Korotkov, A.N., Likharev, K.K.: Possible cooling by resonant Fowler–Nordheim emission, Applied Physics Letters **75** (16), 2491–2493 (1999)
86. Semet, V., Binh, V.T., and Tsu, R.: Shaping electron field emission by ultrathin multilayered structure cathodes, Microelectronics J. **39**, 607–616 (2008)
87. Fisher, T.S., Walker, D.G.: Thermal and electrical energy transport and conversion in nanoscale electron field emission processes, Journal of Heat Transfer, Transactions of the ASME **124** (5), 954–962 (2002)
88. Fisher, T.S., Walker, D.G., Weller, R.A.: Analysis and simulation of anode heating due to electron field emission, IEEE Transactions on Components & Packaging Technologies **26** (2) 317–323 (2003); Walker, D.G., Harris, C.T., Fisher, T.S., Davidson, J.L.: Estimation of parameters in thermal-field emission from diamond, Diamond and Related Materials **14** (1), 113–120 (2005)
89. Tzeng, Y., Chen, Y., Liu, C.: Fabrication and characterization of non-planar high-current-density carbon-nanotube coated cold cathodes, Diamond and Related Materials **12** (3–7), 442–445 (2003)
90. Thong, J.T.L., Oon, C.H., Eng, W.K., Zhang, W.D., and Gan, L.M.: High-current field emission from a vertically aligned carbon nanotube field emitter array, Applied Physics Letters **79** (17), 2811–2813 (2001)
91. Kock, F.A.M., Garguilo, J.M., Brown, B., Nemanich, R.J.: Enhanced low-temperature thermionic field emission from surface-treated N-doped diamond films, Diamond and Related Materials **11** (3–6), 774–779 (2002)
92. Kock, F.A.M., Garguilo, J.M., Wang, Y., Shields, B., and Nemanich, R.J.: Vacuum thermionic emission and energy conversion from nanostructured carbon materials. Materials Research Society Fall Meeting, Symposium S, Boston MA (2003)
93. Koeck, F.A.M., Nemanich, R.J.: Sulfur-doped nanocrystalline diamond films as field enhancement based thermionic emitters and their role in energy conversion, Diamond and

Related Materials **14** (11–12), 2051–2054 (2005); Koeck F.A.M., Nemanich, R.J.: Emission characterization from nitrogen-doped diamond with respect to energy conversion, Diamond and Related Materials **15** (2–3), 217–220 (2006)
94. Moyzhes, B.: Possible ways for efficiency improvement of thermoelectric materials, Fifteenth International Conference on Thermoelectrics, Pasadena, p. 183 (1996)
95. Rowe, D.M., and Min, G.: Multiple potential barriers as a possible mechanism to increase the Seebeck coefficient and electrical power factor, Thirteenth International Conference on Thermoelectrics, Kansas City, pp. 339–342 (1994)
96. Whitlow, L.W., and Hirano, T.: Superlattice applications to thermoelectricity, J. Appl. Phys. **78**, 5460 (1995)
97. Moyzhes, B.Y., and Nemchinsky, V.A.: Report on the XI International Conference on Thermoelectrics, IEEE, Arlington, TX, pp. 232–235 (1992)
98. Shakouri, A., Lee, E.Y., Smith, D.L., Narayanamurti, V., and Bowers, J.E.: Thermoelectric effects in submicron heterostructure barriers, Microscale Thermophysical Eng. **2**, 37–42 (1998)
99. Fan, X.F., Zeng, G.H., LaBounty, C., Bowers, J.E., Croke, E., Ahn, C.C., Huxtable, S., Majumdar, A., and Shakouri, A.: SiGeC/Si superlattice microcoolers, Applied Physics Letters **78**, 1580–1582 (2001)
100. LaBounty, C., Shakouri, A., Abraham, P., Bowers, J.E.: Monolithic integration of thin-film coolers with optoelectronic devices, Optical Engineering **39**, 2847–2852 (2000)
101. Shakouri, A., LaBounty C., Piprek, J., Abraham, P., Bowers, J.E.: Thermionic emission cooling in single barrier heterostructures, Applied Physics Letters **74** (1), 88–89 (1999)
102. LaBounty, C., Shakouri, A., Bowers, J.E.: Design and characterization of thin film microcoolers, Journal of Applied Physics **89** (7), 4059–4064 (2001)
103. Zhang, Y., ,Zeng, G.H., Piprek, J., Bar-Cohen, A., Shakouri, A.: Superlattice microrefrigerators fusion bonded with optoelectronic devices, IEEE Transactions on Components and Packaging Technologies **28** (4), 658–666 (2005)
104. Zebarjadi, M., Esfarjani, K., and Shakouri, A.: Nonlinear Peltier effect in semiconductors, Appl. Phys. Lett. **91**, 122104 (2007)
105. Radtke, R.J., Ehrenreich, H., and Grein, C.H.: Multilayer thermoelectric refrigeration in $Hg_{1-x}Cd_xTe$ superlattices, J. Applied Physics **86**, 3195 (1999)
106. Ulrich, M.D., Barnes, P.A., and Vining, C.B.: Comparison of solid-state thermionic refrigeration with thermoelectric refrigeration, Journal of Applied Physics **90** (3), 1625–1631 (2001)
107. Mahan, G.D., and Vining, C.B.: The B factor in multilayer thermionic refrigeration, J. Appl. Phys. **86**, 6852 (1999)
108. Nolas, G.S., Sharp, J.W., and Goldsmid, H.J.: *Thermoelectrics: Basic Principles and New Material Developments*, Springer-Verlag (2001)
109. Vashaee, D., and Shakouri, A.: Improved thermoelectric power factor in metal-based superlattices, Physical Review Letters **92** (10), 106103–1
110. Shakouri, A., and LaBounty, C.: Material optimization for heterostructure integrated thermionic coolers. Eighteenth International Conference on Thermoelectrics. Proceedings, ICT'99 (Cat. No. 99TH8407), IEEE. pp. 35–9. Piscataway, NJ, USA (1999)
111. O'Dwyer, M.F., Humphrey, T.E., Lewis, R.A., and Zhang, C.: The effect of the electron energy spectrum on electronic efficiency and power in thermionic and thermoelectric devices, 24th International Conference on Thermoelectrics, Clemson, South Carolina, USA (2005)
112. Humphrey, T.E., and O'Dwyer, M.F.: A comparison between solid-state thermionics and thermoelectrics, Journal of Applied Physics **98**, 026108 (2005)
113. Bian, Z., Shakouri, A.: Monte Carlo simulation of solid-state thermionic energy conversion devices based on non-planar heterostructure interfaces, 14th International Conference on Nonequilibrium Carrier Dynamics in Semiconductors (HCIS-14) Chicago, July 24–29 (2005)
114. Zide, J.M., Klenov, D.O., Stemmer, S., Gossard, A.C., Zeng, G., Bowers, J.E., Vashaee, D., and Shakouri, A.: Thermoelectric power factor in semiconductors with buried epitaxial semimetallic nanoparticles, Applied Physics Letters **87** (11), 112102 (2005)

115. Humphrey, T.E., O'Dwyer, M.F., Zhang, C., and Lewis, R.A.: Solid-state thermionics and thermoelectrics in the ballistic transport regime, J. Appl. Phys. **89** (2), 026108 (2005)
116. Mingo, N., Stewart, D.A., Broido, D.A., Srivastava, D.: Phonon transmission through defects in carbon nanotubes from first principles, Phys. Rev. B. **77**, 033418 (2008)
117. Humphrey, T.E., Linke, H.: Reversible thermoelectric nanomaterials, Physical Review Letters **94** (9), 0966011–4 (2005)
118. Nolas, G.S., Slack, G.A., Morelli, D.T., Tritt, T.M., and Ehrlich, A.C.: The effect of rare-earth filling on the lattice thermal conductivity of skutterudites, J. Applied Physics **79**, 4002–4008 (1996)
119. Casimir, H.B.G.: Note on the conduction of heat in crystals, Physica **5**, 495–500 (1938)
120. Venkatasubramanian, R.: Thin-film superlattice and quantum-well structures: A new approach to high-performance thermoelectric materials, Naval Res. Rev. **58**, 44–54 (1996)
121. Yao, T.: Thermal properties of AlAs/GaAs superlattices, Applied Physics Letters **51**, 1798–1800 (1987)
122. Chen, G., Tien, C.L., Wu, X., and Smith, J.S.: Measurement of thermal diffusivity of GaAs/AlGaAs thin-film structures, J. Heat Transfer **116**, 325–331 (1994)
123. Tien, C.L., and Chen, G.: Challenges in microscale conductive and radiative heat transfer, Proc. Am. Soc. Mech. Eng., ASME HTD **227**, 1–12 (1992); also in J. Heat Transfer **116**, 799–807 (1994)
124. Yu, X.Y., Chen, G., Verma, A., and Smith, J.S.: Temperature dependence of thermophysical properties of GaAs/AlAs periodic structure, Applied Physics Letters **67**, 3554–3556 (1995); ibid. **68**, 1303 (1996)
125. Capsinski, W.S., and Maris, H.J.: Thermal conductivity of GaAs/AlAs superlattices, Physica B **220**, 699–701 (1996)
126. Lee, S.M., Cahill, D.G., and Venkatasubramanian, R.: Thermal conductivity of Si–Ge superlattices, Applied Physics Letters **70**, 2957–2959 (1997)
127. Yamasaki, I., Yamanaka, R., Mikami, M., Sonobe, H., Mori, Y., and Sasaki, T.: Thermoelectric properties of Bi_2Te_3/Sb_2Te_3 superlattice structures, Proc. Int. Conf. Thermoelectrics, ICT'98, pp. 210–213 (1998)
128. Capinski, W.S., Maris, H.J., Ruf, T., Cardona, M., Ploog, K., and Katzer, D.S.: Thermal-conductivity measurements of GaAs/AlAs superlattices using a picosecond optical pump-and-probe technique, Physical Review B **59**, 8105–8113 (1999)
129. Borca-Tasciuc, T., Liu, W.L., Zeng, T., Song, D.W., Moore, C.D., Chen, G., Wang, K.L., Goorsky, M.S., Radetic, T., Gronsky, R., Koga, T., and Dresselhaus, M.S.: Thermal conductivity of symmetrically strained Si/Ge superlattices, Superlattices and Microstructures **28**, 119–206 (2000)
130. Song, D.W., Liu, W.L., Zeng, T., Borca-Tasciuc, T., Chen, G., Caylor, C., and Sands, T.D.: Thermal conductivity of skutterudite thin films and superlattices, Applied Physics Letters **77**, 3854–3856 (2000)
131. Venkatasubramanian, R.: Lattice thermal conductivity reduction and phonon localization-like behavior in superlattice structures, Physical Review B **61**, 3091–3097 (2000)
132. Huxtable, S.T., Shakouri, A., LaBounty, C., Fan, X., Abraham, P., Chiu, Y.J., Bowers, and Majumdar, A.: Thermal conductivity of indium phosphide-based superlattices, Microscale Thermophysical Engineering **4**, 197–203 (2000)
133. Borca-Tasciuc, T., Achimov, D., Liu, W.L., Chen, G., Ren, H.-W., Lin, C.-H., and Pei, S.S.: Thermal conductivity of InAs/AlSb superlattices, Microscale Thermophysical Engineering **5**, 225–231 (2001)
134. Liu, W.L., Borca-Tasciuc, T., Chen, G., Liu, J.L., and Wang, K.L.: Anisotropy thermal conductivity of Ge quantum dot and symmetrically strained Si/Ge superlattices, J. Nanoscience and Nanotechnology **1**, 39–42 (2001)
135. Narayanamurti, V., Stormer, J.L., Chin, M.A., Gossard, A.C., and Wiegmann: Selective transmission of high-frequency phonons by a superlattice: The 'dielectric' phonon filter, Physical Review Letters **43**, 2012–2015 (1979)
136. Ren, S.Y., and Dow, J.: Thermal conductivity of superlattices, Physical Review B **25**, 3750–3755 (1982)

137. Chen, G.: Size and interface effects on thermal conductivity of superlattices and periodic thin-film structures, J. Heat Transfer **119**, 220–229 (1997); see also Proc. Nat. Heat Transf. Conf., ASME HTD **323**, 121 (1996)
138. Hyldgaard, P., and Mahan, G.D.: Phonon Knudson flow in superlattices, Thermal Conductivity **23**, 172–181, Technomic, Lancaster (1996)
139. Chen, G., and Neagu, M.: Thermal conductivity and heat transfer in superlattices, Applied Physics Letters **71**, 2761–2763 (1997)
140. Chen, G.: Thermal conductivity and ballistic phonon transport in cross-plane direction of superlattices, Physical Review B **57**, 14958–14973 (1998); see also Proc. Int. Mech. Eng. Congr., DSC **59**, 13 (1996)
141. Hyldgaard, P. and Mahan, G.D.: Phonon superlattice transport, Physical Review B **56**, 10754–10757 (1997)
142. Tamura, S., Tanaka, Y., and Maris, H.J.: Phonon group velocity and thermal conduction in superlattices, Physical Review B **60**, 2627–2630 (1999)
143. Bies, W.E., Radtke, R.J., Ehrenreich, H.: Phonon dispersion effects and the thermal conductivity reduction in GaAs/AlAs superlattices, J. Applied Physics **88**, 1498–1503 (2000)
144. Kiselev, A.A., Kim, K.W., Stroscio, M.A.: Thermal conductivity of Si/Ge superlattices: A realistic model with a diatomic unit cell, Physical Review B **62**, 6896–6899 (2000)
145. Yang, B., and Chen, G.: Lattice dynamics study of anisotropic heat conduction in superlattices, Microscale Thermophysical Engineering **5** (2) 107–116 (10)(2001)
146. Volz, S.G., Saulnier, J.B., Chen, G., and Beauchamp, P.: Computation of thermal conductivity of Si/Ge superlattices by molecular dynanmics techniques, Microelectronics J. **31**, 815–819 (2000)
147. Liang, X.G. and Shi, B.: Two-dimensional molecular dynamics simulation of the thermal conductance of superlattices with Lennard-Jones potential, Materials Science and Engineering **292**, 198–202 (2000)
148. Chen, G.: Phonon wave heat conduction in thin films and superlattices, J. Heat Transfer **121**, 945–953 (1999)
149. Ju, Y.S., and Goodson, K.E.: Phonon scattering in silicon films with thickness of order 100 nm, Applied Physics Letters **74**, 3005–3007 (1999)
150. Ziman, J.M.: *Electrons and Phonons*, Clarendon Press, Oxford (1960)
151. Venkatasubramian, R., Siivola, E., and Colpitts, T.S.: In-plane thermoelectric properties of freestanding Si/Ge superlattice structures, Proc. Int. Conf. Thermoelectrics, ICT'98, pp. 191–197 (1998)
152. Simkin, M.V., and Mahan, G.D.: Minimum thermal conductivity of superlattices, Physical Review Letters **84**, 927–930 (2000)
153. Arutyunyan, L.I., Bogomolov, V.N., Kartenko, N.F., Kurdyukov, D.A., Popov, V.V., Prokof'ev, A.V., Smirnov, I.A., and Sharenkova, N.V.: Thermal conductivity of a new type of regular-structure nanocomposites: PbSe in opal pores, Physics of the Solid State **39**, 510–514 (1997)
154. Song, D.W., Shen, W.-N., Zeng, T., Liu, W.L., Chen, G., Dunn, B., Moore, C.D., Goorsky, M.S., Radetic, R., and Gronsky, R.: Thermal conductivity of nanoporous bismuth thin films for thermoelectric applications, ASME HTD **364-1**, 339–344 (1999)
155. Volz, S.G., and Chen, G.: Molecular dynamics simulation of thermal conductivity of Si nanowires, Applied Physics Letters **75**, 2056-2058 (1999)
156. Walkauskas, S.G., Broido, D.A., Kempa, K., and Reinecke, T.L.: Lattice thermal conductivity of wires, J. Applied Physics **85**, 2579–2582 (1999)
157. Shi, L., Li, D.Y., Yu, C.H., Jang, W.Y., Kim, D., Yao, Z., Kim, P., Majumdar, A.: Measuring thermal and thermoelectric properties of one-dimensional nanostructures using a microfabricated device, J. of Heat Transfer, Transactions of the ASME **125** (5), 881–888 (2003)
158. Zhou, F., Seol, J.H., Moore, A.L., Shi, L., Ye, Q., Scheffler, R.: One-dimensional electron transport and thermopower in an individual InSb nanowire, J. Phys.: Condens. Matter **18**, 9651–9657 (2006)
159. Boukai, A.I., Bunimovich, Y., Tahir-Kheli, J., et al.: Silicon nanowires as efficient thermoelectric materials, Nature **451**, 168 (2008)

160. Shi, L.: The role of defects on thermal and thermoelectric Transport in nanowires, MRS Spring Meeting, San Francisco (2009)
161. Savic, I., Mingo, N., and Stewart, D.A.: Phonon transport in isotope-disordered carbon and boron-nitride nanotubes: Is localization observable? Phys. Rev. Lett. **101**, 165502 (2008); For a different explanation of the possible low thermal conductivity see: Martin, P., Aksamija, Z., Pop, E., et al.: Impact of phonon-surface roughness scattering on thermal conductivity of thin Si nanowires, Phys. Rev. Lett. **102**, 125503 (2009)
162. Shakouri, A.: Nanoscale thermal transport and microrefrigerators on a chip, Proc. of IEEE **94**, no. 8, 1613–1638 (2006)
163. Singh, R., Vashaee, D., Zhang, Y., Negassi, M., Shakouri, A., Okuno, Y., Zeng, G., LaBounty, C., and Bowers, J.E.: Experimental characterization and modeling of InP-based microcoolers, Materials Research Society Fall Meeting, S11.4, Boston, MA (2003)
164. LaBounty, C.J., Shakouri, A., Robinson, G., Esparza, L., Abraham, P., and Bowers, J.E.: Experimental investigation of thin film InGaAsP coolers. Thermoelectric Materials. The Next Generation Materials for Small-Scale Refrigeration and Power Generation Applications, Materials Research Society Symposium Proceedings **626**, Z14.4.1–6 (2001)
165. Zhang, Y., Vashaee, D., Singh, R., Shakouri, A., Zeng, G. and Chiu, Y.-J.: Influence of doping concentration and ambient temperature on the cross-plane Seebeck coefficient of InGaAs/InAlAs superlattices. Thermoelectric Materials 2003. Research and Applications Symposium, Materials Research Society Symposium Proceedings **793**, 59–65 (2004)
166. LaBounty, C., Almuneau, G., Shakouri, A., Bowers, J.E.: Sb-based III-V cooler. 19th International Conference on Thermoelectrics, Cardiff, Wales (2000)
167. Zeng, G., Shakouri, A., Bounty, C.L., Robinson, G., Croke, E., Abraham, P., Fan, X., Reese, H., Bowers, J.E.: SiGe micro-cooler, Electronics Letters **35** (24), 2146–7, 25 (1999)
168. Fan, X., Zeng, G., LaBounty, C., Croke, E., Vashaee, D., Shakouri, A., Ahn, C., Bowers, J.E.: High cooling power density SiGe/Si microcoolers, Elec. Lett. **37** (2), 126 (2001)
169. Fan, X., Zeng, G., Croke, E., Robinson, G., LaBounty, C., Shakouri, A., and Bowers, J.E.: SiGe/Si superlattice coolers, Physics of Low-Dimensional Structures **5–6**, 1–9 (2000)
170. Zhang, Y., Zeng, G., Shakouri, A., Wang, P., Yang B. and Bar-Cohen, A.: Optimization of doping concentration for three-dimensional bulk silicon microrefrigerators, 22nd IEEE SEMI-THERM Symposium (2006)
171. Dilhaire, S., Ezzahri, Y., Grauby, S., Claeys, W., Christofferson, J., Zhang, Y. and Shakouri, A.: Thermal and thermomechanical study of micro-refrigerators on a chip based on semiconductor heterostructures. 22nd International Conference on Thermoelectrics (Cat. No.03TH8726), IEEE pp. 519–523 (2003)
172. Zhang Y., Vashaee D., Christofferson, J., Zeng, G., LaBounty, C., Piprek, J., Croke, Ed., and Shakouri, A.: 3D electrothermal simulation of heterostructure thin film microcooler. 2003 ASME Symposium on the Analysis and Applications of Heat Pump & Refrigeration Systems Proceedings, Washington, DC (2003)
173. Vashaee, D., Christofferson, J., Zhang, Y., Shakouri, A., Zeng, G.H., LaBounty, C., Fan, XF., Piprek, J., Bowers, J.E., and Croke., E.: Modeling and optimization of single-element bulk SiGe thin-film coolers, Microscale Thermophysical Engineering **9** (1), 99–118 (2005)
174. Zeng, G., Fan, X., Labounty, C., Croke, E., Zhang, Y., Christofferson, J., Vashaee, D., Shakouri A., and Bowers, J.E.: Cooling power density of SiGe/Si superlattice micro-refrigerators, Materials Research Society Fall Meeting 2003, Proceedings **793**, 43 (2003)
175. Shakouri, A., Labounty C.: Material optimization for heterostructure integrated thermionic coolers, International Conference on Thermoelectrics, Baltimore, MD (1999)
176. Fitting, A., Christofferson, J., Shakouri, A., Fan, X., Zeng G., LaBounty, C., Bowers, JE., Croke, E.: Transient response of thin film SiGe microcoolers, International Mechanical Engineering Congress and Exhibition (IMECE 2001), New York, NY (2001)
177. Shakouri, A., Zhang, Y.: On-chip solid-state cooling for integrated circuits using thin-film microrefrigerators, IEEE Transactions on Components and Packaging Technologies, **28** (1), 65–69 (2005); Zhang Y., Christofferson, J., Shakouri, A., Zeng, G., Bowers, J.E., Croke, E.: High-speed localized cooling using SiGe superlattice microrefrigerators. Nineteenth Annual IEEE Semiconductor Thermal Measurement and Management Symposium (Cat. No. 03CH37437), pp. 61–65 (2003)

178. Fukutani, K., Zhang, Y. and Shakouri, A.: Solid-state microrefrigerators on a chip, Electronic Cooling **12** (3), 18–26 (2006)
179. Fan, X.: Ph.D. dissertation, Univ. California, Santa Barbara
180. Vashaee, D.: Ph.D. dissertation, Univ. California, Santa Cruz
181. Christofferson, J., and Shakouri, A.: Thermoreflectance based thermal microscope, Rev. Sci. Instrum. **76**, 024903–1 (2005)
182. Huxtable, S.T., Abramson, A.R., Tien, C.-L., Majumdar, A., LaBounty, C., Fan, X., Zeng, G., Bowers, J.E., Shakouri, A., Croke, E.T.: Thermal conductivity of Si/SiGe and SiGe/SiGe superlattices, Applied Physics Letters **80** (10), 1737–1739 (2002); Li, D., Huxtable, S.T., Abramson, A.R., Majumdar, A.: Thermal transport in nanostructured solid-state cooling devices, Transactions of the ASME, Journal of Heat Transfer **127** (1), 108–114 (2005)
183. Cahill D.G.: Thermal conductivity measurement from 30 to 750 K: The 3 omega method, Rev. Sci. Instrum. **61**, 802 (1990)
184. Fan, X., Zeng, G., LaBounty, C., Vashaee, D., Christofferson, J., Shakouri, A., and Bowers, J.E.: Integrated cooling for Si-based microelectronics. 20th International Conference on Thermoelectrics (Cat. No. 01TH8589), Beijing, China, 405–408 (2001)
185. Ezzahri, Y., Singh, R., Christofferson, J., Bian, Z., Shakouri, A.: Optimization of Si/SiGe microrefrigerators for hybrid solid-state/liquid cooling, InterPACK, paper 33878 (2007)
186. Ezzahri, Y., Zeng, G., Fukutani, K., et al.: A comparison of thin film microrefrigerators based on Si/SiGe superlattice and bulk SiGe, Microelectronics J. **39**, 981 (2008)
187. Rawat, V., and Sands, T.: Growth of TiN/GaN metal/semiconductor multilayers by reactive pulsed laser deposition, J. Appl. Phys. **100**, 064901 (2006); Rawat, V., and Sands, T.: TiN/GaN metal/semiconductor multilayer nanocomposites grown by reactive pulsed laser deposition, MRS Proc. **872**, J21.4.1–6 (2005)
188. Kim, W., Singer, S., Majumdar, A., Zide, J., Gossard, A., and Shakouri, A.: Role of nanostructures in reducing thermal conductivity below alloy limit in crystalline solids, 24th International Conference on Thermoelectrics, pp. 9–12 (2005)
189. Kim, W., Singer, S., Majumdar, A., Vashaee, D., Bian, Z., Shakouri, A., Zeng, G., Bowers, J.E., Zide, J.M.O., Gossard, A.C.: Cross-plane lattice and electronic thermal conductivities of ErAs: InGaAs/InGaAlAs superlattices, Appl. Phys. Lett. **88**, 242107 (2006)
190. Kim, W., Zide, J.M., Gossard, A., Klenov, D., Stemmer, S., Shakouri, A., Majumdar, A.: Thermal conductivity reduction and thermoelectric figure of merit increase by embedding nanoparticles in crystalline semiconductors, Phys. Rev. Lett. **96**, no. 4, 045901/1–4 (2006)
191. Zeng, G., Zide, J.M.O., Kim, W., Bowers, J.E., Gossard, A.C., Bian, Z., Zhang, Y., Shakouri, A., Singer, S.L., Majumdar, A.: Cross-plane Seebeck coefficient of ErAs:InGaAs/InGaAlAs superlattices, Journal of Applied Physics **101** (3), 34502-1–5 (2007)
192. Zide, J.M., Vashaee, D., Zeng, G., Bowers, J.E., Shakouri, A., and Gossard, A.C.: Demonstration of electron filtering to increase the Seebeck coefficient in ErAs:InGaAs/InGaAlAs superlattices, Physical Review B **74** (20), 205335-1–5 (2006)
193. Zebarjadi, M., Esfarjani, K., Shakouri, A., Bahk, JH., Bian, ZX., Zeng, G., Bowers, J.E., Lu, H., Zide, J.M.O., and Gossard, A.: Effect of nanoparticle scattering on thermoelectric power factor, to appear in App. Phys. Lett. (2009)
194. Zebarjadi, M., Esfarjani, K., Shakouri, A., Bian, Z.X., Bahk, J.H., Zeng, G., Bowers, J.E., Lu, H., Zide, J.M.O., and Gossard, A.: Effect of nanoparticle scattering on electronic and thermoelectric transport, J. of Electronic Materials, first published online (DOI: 10.1007/s11664-008-0656-4) (2009)
195. Zeng, G., Bahk, J.-H., Bowers, J.E., Zide, J.M.O., Gossard, A.C., Bian, Z., Singh, R., Shakouri, A., Kim, W., Singer, S.L., and Majumdar, A.: ErAs:$(InGaAs)_{1-x}(InAlAs)_x$ alloy power generator modules, Appl. Phys. Lett. **91**, 263510 (2007)
196. Rawat, V., Koh, Y.K., Cahill, D.G, et al.: Thermal conductivity of (Zr,W)N/ScN metal/semiconductor multilayers and superlattices, J. Appl. Phys **105**, 024909 (2009)
197. Zebarjadi, M., Bian, Z.X., Singh, R., Shakouri, A., Wortman, R., Rawat, V., and Sands, T.: Thermoelectric transport in a ZrN/ScN Superlattice, Journal of Electronic Materials, first published online (DOI: 10.1007/s11664-008-0639-5) (2009)

198. Wilk, A., Kovsh, A.R., Mikhrin, S.S., Chaix, C., Novikov, I.I., Maximov, M.V., Shernyakov, Yu.M., Ustinov, V.M., Ledentsov, N.N.: High-power 1.3 μm InAs/GaInAs/GaAs QD lasers grown in a multiwafer MBE production system, Journal of Crystal Growth **278**, No. 1–4, 335–341 (2005); Bacher, K., Massie, S., Hartzel, D., Stewart, T.: Present ability of commercial molecular beam epitaxy, International Conference on Indium Phosphide and Related Materials (Cat. No. 97CH36058), pp. 351–352 (1997)
199. Zeng, G., Bowers, J.E., Zide, J.M.O., Gossard, A.C., Kim, W., Singer, S., Majumdar, A., Singh, R., Bian, Z., Zhang, Y., and Shakouri, A.: ErAs:InGaAs/InGaAlAs superlattice thin-film power generator array, Applied Physics Letters **88**, 13502-1–3 (2006)
200. Dutta, N.K., Cella, T., Brown, R.L., Huo, D.T.C.: Monolithically integrated thermoelectric controlled laser diode, Applied Physics Letters **47**, 222–224 (1985)
201. Chen, G.: Heat transfer in micro- and nanoscale photonic devices, Annual Review of Heat Transfer **7**, 1–57 (1996)
202. Piprek, J., Akulova, Y.A., Babifc, D.I., Coldren, L.A., Bowers, J.E.: Minimum temperature sensitivity of 1.55 μm vertical-cavity lasers at 30 nm gain offset, Applied Physics Letters **72**, 1814–1816 (1998)
203. Berger, P.R., Dutta, N.K., Choquette, K.D., Hasnain, G., Chand, N.: Monolithically Peltier-cooled vertical-cavity surface-emitting lasers, Applied Physics Letters **59**, 117–119 (1991)
204. Corser, T.A.: Qualification and reliability of thermoelectric coolers for use in laser modules. 41st Electronic Components and Technology Conference, Atlanta, GA, USA, May, pp. 150–156 (1991)
205. Cheng, Y.K., Tsai, C.H., Teng, C.C., and Kang, S.M.: *Electrothermal Analysis of VLSI Systems*, Kluwer Academic Publishers, Dordrecht (2000)
206. `http://www.thermion-company.com`
207. `http://www.micropelt.com/`
208. `http://www.nextremethermal.com`
209. Chowdhury, Prasher R., Lofgreen, K., Chrysler, G., Narasimhan, S., Mahajan, R., Koester, R., Alley, R., and Venkatasubramanian, R.: On-chip cooling by superlattice-based thin-film thermoelectrics, Nature Nanotechnology **4**, 235–238 (2008)
210. Kishi, M., Nemoto, H., Hamao, T., Yamamoto, M., Sudou, S., Mandai, M., and Yamamoto, S.: Micro thermoelectric modules and their application to wrist watches as an energy source, Proc. Int. Conf. Thermoelectrics, pp. 301–307 (1999)
211. Leonov, V., Torfs, T., Fiorini, P., Van Hoof, C.: Thermoelectric converters of human warmth for self-powered wireless sensor nodes, IEEE Sensors Journal **7**, 650–657 (2007)
212. Muller, E., Walczak, S., and Seifert, W.: Optimization strategies for segmented Peltier coolers, Phys. Stat. Sol. (a) **203**, 2128 (2006)
213. Snyder, G.J., and Ursell, T.S.: Thermoelectric efficiency and compatibility, Phys. Rev. Lett. **91**, 148301 (2003)
214. Bian, Z., and Shakouri, A.: Beating the maximum cooling limit with graded thermoelectric materials, Appl. Phys. Lett. **89**, 212101 (2006)
215. Bian, Z., Wang, H., Zhou, Q., and Shakouri, A.: Maximum cooling temperature and uniform efficiency criterion for inhomogeneous thermoelectric materials, Phys. Rev. B **75**, 245208 (2007)
216. Lee, S.M., and Cahill, D.G.: Heat transport in thin dielectric films, J. Applied Physics **81**, 2590–2595 (1997)
217. Borca-Tasciuc, T., Kumar, R., and Chen, G.: Data reduction in 3ω method for thin film thermal conductivity measurements, Review of Scientific Instruments **72**, 2139–2147 (2001)
218. Cahill, D.G., Fischer, H.E., Klitsner, T., Swartz, E.T., and Pohl, R.O.: Thermal conductivity of thin films: Measurement and understanding, J. Vacuum Science and Technolnology A **7**, 1259–1266 (1989)
219. Hatta, I.: Thermal diffusivity measurement of thin films and multilayered composites, Int. J. Thermophys. **11**, 293–303 (1990)
220. Volklein, F., and Starz, T.: Thermal conductivity of thin films: Experimental methods and theoretical interpretation, Proc. Int. Conf. Thermoelectrics, ICT'97, pp. 711–718 (1997)

221. Goodson, K.E., and Ju, Y.S.: Heat conduction in novel electronic films, Ann. Rev. Mat. **29**, 261–293 (1999)
222. Cahill, D.G.: Analysis of heat flow in layered structures for time-domain thermoreflectance, Rev. Sci. Instrum. **75**, 5119 (2004)
223. Schmidt, A.J., Chen, X., and Chen, G.: Pulse accumulation, radial heat conduction, and anisotropic thermal conductivity in pump–probe transient thermoreflectance, Review of Scientific Instruments **79**, 114902 (2008)
224. Huxtable, S., Cahill, D.G., Fauconnier, V., White, J.O., and Zhao, J.C.: Thermal conductivity imaging at micron-scale resolution for combinatorial studies of materials, Nature Materials **3**, 298–301 (2004)
225. Yang, B., Liu, J.L., Wang, K.L., and Chen, G.: Simultaneous measurements of Seebeck coefficient and thermal conductivity across superlattice, Applied Physics Letters **80**, 1758–1760 (2002)
226. Zeng, G., Bowers, J.E., Zhang, Y., Shakouri, A., Zide, J., Gossard, A., Kim, W., and Majumdar, A.: ErAs/InGaAs superlattice Seebeck coefficient, Proceedings of the 24th International conference on Thermoelectrics, Clemson, SC, pp. 485–488 (2005)
227. Singh, R., Bian, Z., Zeng, G., Zide, J., Christofferson, J., Chou, H., Gossard, A., Bowers, J.E., and Shakouri, A.: Transient Harman measurement of the cross-plane *ZT* of InGaAs/InGaAlAs superlattices with embedded ErAs nanoparticles, Proceedings of MRS Fall Meeting, Boston (2005)
228. Bian, Z., Zhang, Y., Schmidt, H., Shakouri, A.: Thin film *ZT* characterization using transient Harman technique, 24th International Conference on Thermoelectrics (ICT) (Cat. No.05TH8854C), IEEE, pp. 76–78 (2005)
229. Harman, T.C.: Special techniques for measurement of thermoelectric properties, Journal of Applied Physics **29** (9), 1373–1374 (1958)
230. Singh, R., and Shakouri, A.: Thermostat for high temperature and transient characterization of thin film thermoelectric materials, Rev. of Scientific Instruments **80**, 025101 (2009); For a detailed analysis of the cross-plane *ZT* and Seebeck extraction see: Singh, R., et al.: to be published in Applied Physics Letters (2009)
231. Mavrokefalos, A., Pettes, M.A., Zhou, F., Shi, L.: Four-probe measurements of the in-plane thermoelectric properties of nanofilms, Review of Scientific Instruments **78**, 034901 (2007)
232. R.E. Nelson: A brief history of thermophotovoltaic development, Semiconductor Science and Technology **18** (5), S141–S143 (2003); Harder, N.-P., Wurfel, P.: Theoretical limits of thermophotovoltaic solar energy conversion, Semiconductor Science and Technology **18** (5), S151–S157 (2003)
233. Green, M.: *Third Generation Photovoltaics and Advanced Solar Conversion*, Springer-Verlag (2003)
234. Baldasaro, P.F., Dashiell, M.W., Oppenlander, J.E., Vell, J.L., Fourspring, P., Rahner, K., Danielson, L.R., Burger, S., Brown, E.: System performance projections for TPV energy conversion, American Institute of Physics Conference Proceedings, no. 738, pp. 61–70 (2004); Baldasaro, P.F., Raynolds, J.E., Charache, G.W., DePoy, D.M., Ballinger, C.T., Donovan, T., Borrego, J.M.: Thermodynamic analysis of thermophotovoltaic efficiency and power density tradeoffs, Journal of Applied Physics **89** (6), 3319–3327 (2001)
235. Humphrey, T.E., and Linke, H.: Inhomogeneous doping in thermoelectric nanomaterials. Plenary presentation at the International Thermoelectrics Conference, Adelaide (2004) `cond-mat/0407506`
236. Kittel, K., and Kroemer, H.: *Thermal Physics*, 2nd edn., W.H. Freeman Company (1980)
237. Sander, M.S., Gronsky, R., Sands, T., Stacy, A.M.: Structure of bismuth telluride nanowire arrays fabricated by electrodeposition into porous anodic alumina templates, Chemistry of Materials **15** (1), 335–339 (2003)
238. Radtke, R.J., Ehrenreich, H., and Grein, C.H.: Multilayer thermoelectric refrigeration in $Hg_{1-x}Cd_xTe$ superlattices, Journal of Applied Physics **86** (6) 3195–3198 (1999)
239. Zeng, T.F., and Chen, G.: Energy conversion in heterostructures for thermionic cooling, Microscale Thermophysical Engineering **4**, 39–50 (2000)

240. Reggiani, L. (Ed.): *Hot-Electron Transport in Semiconductors*, Springer-Verlag (1985)
241. Qiu, T.Q., and Tien, C.L.: Heat transfer mechanisms during short-pulse laser heating of metals, J. Heat Transfer **115**, 835–841 (1993)
242. Zeng, T.F., and Chen, G.: Nonequilibrium electron and phonon transport in heterostructures for energy conversion, Proceedings of Int. Mech. Eng. Congress and Exhibition (IMECE2000), ASME HTD, Vol. **366-2**, 361–372 (2000)
243. Vashaee, D. and Shakouri, A.: Non-equilibrium electrons and phonons in heterostructure integrated thermionic coolers, Microscale Thermophysical Engineering **8**, 91–100 (2004)
244. Sze, S.M.: *Physics of Semiconductor Devices*, 2nd edn., Wiley Interscience (1981)
245. Pipe, K.P., Ram, R.J., and Shakouri, A.: Bias-dependent Peltier coefficient and internal cooling in bipolar devices, Phys. Rev. B **66**, 125316, 27, 1–11 (2002)
246. Lyeo, H.K., Khajetoorians, A.A., Shi, L., Pipe, K.P., Ram, R.J., Shakouri, A., and Shih, C.K.: Profiling the thermoelectric power of semiconductor junctions with nanometer resolution, Science **303**, 816–818 (2004)
247. Shakouri, A., Bowers, J.E.: Heterostructure integrated thermionic refrigeration, 16th International Conference on Thermoelectrics (Cat. No.97TH8291), IEEE, pp. 636–640 (1997)
248. Pipe, K.P., Ram, R.J., and Shakouri, A.: Internal cooling in a semiconductor laser diode, IEEE Photonics Technology Letters **14** (4), 453–455 (2002)
249. Mungan, C.E., Buchwald, M.I., Edwards, B.C., Epstein, R.I., and Gosnell, T.R.: Laser cooling of a solid by 16 K starting from room temperature, Physical Review Letters **78** (6), 1030 (1997)

Chapter 10
Molecular Probes for Thermometry in Microfluidic Devices

Charlie Gosse, Christian Bergaud, and Peter Löw

The temperature is an important parameter with regard to chemical reactivity. It is therefore essential to ensure good thermal control within microsystems designed to carry out biological analysis. We begin by reviewing temperature measurement in the context of the lab-on-a-chip, and outlining the various generic strategies available. We then turn more specifically to luminescent molecular probes. We shall show that they all exploit the effect of temperature on a chemical reaction (in the broad sense of the term). More precisely, these probes can be divided in three main categories depending on whether one relies on a phase transition, the modification of a reaction rate, or a shift in an equilibrium. We shall also discuss the main experimental strategies used to transform the image obtained by fluorescence microscopy into a thermal map. Finally, we shall extend the discussion to a few other spectroscopic techniques and examine the prospects for this particular area of microfluidics.

10.1 Microlaboratories and Heat Transfer Issues

10.1.1 Historical Context

For the present discussion we shall take a lab-on-a-chip (LOC) [1, 2] to be a microsystem designed for chemical and biological analysis, with a size of a few square centimeters, able to output the desired data in digital form when supplied with a raw sample. This field of research came into being some fifteen years ago, the offspring of four parents: the defence industry, molecular biology, microelectronics, and analytical sciences [1]. The first two provided the motivation. Indeed, the emergence of biological threats made it essential to develop portable devices soldiers could use to gather information about their environment. At the same time, large-scale sequencing projects were bringing biology into a new age, where mass data acquisition would be of fundamental importance. Automation was the

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 301
DOI 10.1007/978-3-642-04258-4_10, © Springer-Verlag Berlin Heidelberg 2009

solution chosen to achieve higher throughput and better reproducibility. At this point, the concepts of integration and parallelisation could be adopted from microelectronics: the various biotechnological operations would be carried out on a single chip, where microfluidic channels would serve to transport the sample from one active zone to the next [3] (similarly to the electrical tracks connecting the different components on a computer board). Since miniaturisation had thus become implicitly inescapable, scientists began to exploit microfabrication techniques previously developed for integrated circuits and microelectromechanical systems (MEMS) [2]. Finally, this general trend also benefited from advances recently made in analytical sciences: advantageous effects were coming to light in the context of scale reduction, with important progress in terms of resolution, sensitivity, and separation time [3]. Starting with a relatively simple device for gas phase chromatography [4], the field of on-chip analysis began to witness the development of ever more complex and highly integrated systems [5]. Today, we can diagnose the presence of anthrax in less than 30 minutes and from a mere 750 nL of blood [6].

In Sects. 10.1.2–10.1.4, we shall discuss in more detail how the miniaturisation of analysis systems has provided a way round a certain number of technological bottlenecks related to heat transfer issues.

10.1.2 Electrokinetic Separation

Under the influence of an electric field $\mathbf{E}$, any charged molecule in solution moves with velocity $\mathbf{U} = \mu\mathbf{E}$. Electrophoresis can thus separate compounds according to their electrophoretic mobility μ. Historically, migration was performed in a thin layer of aqueous gel ($\sim$ 0.2–5 mm). However, it was soon realised that the current flow caused non-negligible heating of the buffer, and therefore had harmful effects on the quality of the analysis. This problem was in part solved by working with very fine capillary tubes ($\sim$ 50–150 μm), then microchannels ($\sim$ 10 μm), where the increased surface to volume ratio allows better dissipation of heat [7, 8]. To be more precise, in such capillaries the temperature rise can be calculated at any point of the liquid cylinder of radius R by solving the heat diffusion equation. Assuming that the power dissipated per unit volume is homogeneous and equal to σE^2, this leads to

$$T(r) \approx T_{\text{ext}} + \frac{\sigma E^2}{2}\left(\frac{R^2 - r^2}{2k} + \alpha R^2 + \beta R\right), \tag{10.1}$$

where r is the distance to the tube axis, T_{ext} is the temperature of the surrounding medium, σ is the electrical conductivity of the buffer, and k its thermal conductivity. Additionally, α and β are numerical factors accounting for the geometrical and thermal properties of the capillary system (including the buffer).

The temperature can increase by more than ten degrees when cooling depends solely on natural convection in the surrounding air. Many physicochemical

parameters, e.g., pH, molecular interactions, stability of the biological material, are then modified in ways that are not necessarily very well controlled or understood (see Sect. 10.5). Moreover, since the conductivity of the buffer increases with temperature ($\sim 1.9 \times 10^{-2}\,\mathrm{K}^{-1}$), the Joule effect will be strengthened and this positive feedback can in some cases lead to the evaporation of the separation medium [9]. Finally, note also the appearance of a parabolic temperature profile in (10.1). This leads to peak broadening by Taylor–Aris dispersion. Indeed, since the buffer is hotter at the capillary axis, it will be less viscous, whence molecules will be able to migrate more quickly there than near the walls.

In order to characterise the separation more quantitatively, one defines the plate height H as the distance above which displacements by migration dominate over displacements by diffusion, whence $H = 2D/U$ with D the diffusion coefficient of the considered molecule. Since the resolution of an analysis increases with the narrowness of the peaks, its effectiveness is characterised by the number N of theoretical plates contained in the capillary of length L:

$$N = \frac{L}{H} = \frac{\mu}{2D} LE \,. \tag{10.2}$$

Independently, the analysis rate is the other parameter to be taken into account when developing a separation technique:

$$\tau_{\text{analysis}} = \frac{L}{\mu E} \,. \tag{10.3}$$

To obtain a faster and higher quality separation, it will thus be worth increasing the electric field: $E \to aE$ [see (10.2) and (10.3)]. In fact, this will only be possible at constant temperature, i.e., without additional dispersion, if the capillary radius is divided by the same factor, viz., $R \to R/a$ [see (10.1)]. It is this increased throughput and resolution which explains in part the success of capillary electrophoresis, as well as the final boost that had characterized the human genome sequencing project [10].

Finally, note that on-chip separation columns differ from capillaries both by their smaller diameter and by a reduction in their length. Starting with a macroscopic device, they have thus undergone the two transformations $L \longrightarrow L/a$ and $R \longrightarrow R/a$. So an increased field will only allow a higher analysis rate [see (10.3)] at constant quality [see (10.1) and (10.2)]. Separation times of less than a millisecond can now be achieved with L of the order of a few hundred microns [11].

10.1.3 DNA Amplification by PCR

The polymerase chain reaction (PCR) is an enzymatic reaction providing exponential and selective amplification of a target DNA sequence. Starting with an initially rather scant sample, enough material can then be obtained to be able to carry out

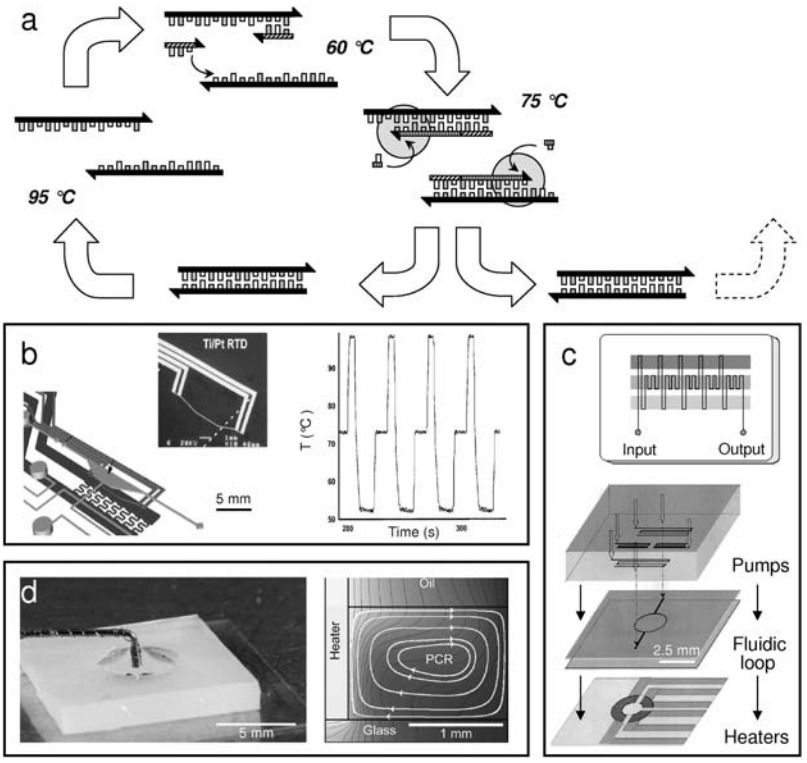

Fig. 10.1 (**a**) Polymerase chain reaction (PCR) principle. The reaction mixture is first heated to about 95°C to denature the DNA double helix serving as substrate. The sample is then cooled to around 60°C in order to hybridise the two primers onto their target. Binding of these short fragments of single strand DNA is sequence specific, so only the desired sequence will be amplified. Finally, the temperature is slightly raised, to around 75°C, so that the DNA polymerase can replicate, from each of the primers, the template on which it is fixed. After repeating this thermal cycle n times, it yields about 2^n copies of the target sequence. (**b**–**d**) Three different ways of carrying out the temperature cycle required for PCR. (**b**) *Left*: Microdevice where the solution is immobile and temperature variations are obtained by Joule heating. *Right*: Corresponding temperature measurement obtained using a resistive sensor. From [19]. (**c**) Two microsystems where the sample moves through regions at different temperatures. *Top*: Winding channel with continuous injection of reagents. *Bottom*: Closed loop where the reaction mixture is set in motion using an integrated peristaltic pump. Taken from [22, 24]. (**d**) Microfluidic cell where the reagents passes from a hot region to a cold region thanks to convection. From [28]

accurate genetic tests. Given the spin-offs in pathogen detection, diagnosis, and forensic medecine, it soon became essential to integrate this technique into microlaboratories [12–14].

Three temperature changes are required at each stage in the amplification process (see Fig. 10.1a), and once again reducing the size of the apparatus leads to a significant improvement in performance. Indeed, for one thing, the cycle frequency

can be increased, because it is usually limited by the rate at which the temperature of the reaction mixture can be varied, itself governed by the thermal inertia of the reactor (including the sample). For another, the enzymatic reaction becomes more efficient and more specific on-chip, because the thermal gradients in the solution are reduced. The first devices, microfabricated in silicon, divided both the volume and the reaction time by a factor of more than two [15, 16]. Many different strategies were then explored, particularly with regard to the type of heating, e.g., macroscopic Peltier module [16, 17], infrared radiation [18], microfabricated resistor [15, 19] (see Fig. 10.1b). Another trend was toward greater integration, with electrophoretic analysis of the PCR product performed on-line [5,6,17,19,20]. At the present time, amplification by a microsystem takes some 10–15 minutes, using a volume of 200–500 nL, while typical magnitudes for a tube reaction are greater than 30 minutes and 10 μL, respectively.

In parallel with the straight miniaturisation of macroscopic apparatus, innovative techniques were also put into practice. For instance, devices were developed in which it is no longer the heating or cooling of the reactor that modifies the temperature of the reaction mixture, but rather the displacement of the latter between regions thermostated at different temperatures. Therefore problems of thermal inertia are reduced to an absolute minimum, i.e., to the sole thermal mass of the PCR solution. Examples include continuous flow devices [21], where the sample moves cyclically over different hot plates, either because it follows an open winding channel [22,23], or because it is pumped through a closed loop (see Fig. 10.1c) [24]. Even more ingeniously, convective reactors remove any need for active displacement of the fluid, the reaction mixture being circulated by virtue of a thermo-hydrodynamic instability (see Fig. 10.1d) [25–28].

10.1.4 Thermodynamic and Kinetic Measurements

Calorimetry is a technique that allows to fully characterise the thermodynamics of molecular interactions. Since it requires no labelling, it is totally generic and hence widely used in drug discovery [29, 30]. Unfortunately, the large amount of sample and the long measurement time required are often prohibitive. Nevertheless, miniaturisation looks today like a good way to decrease the cost and increase the throughput of screening.

Consider the binding of a ligand on a receptor, expressed by

$$\mathrm{L} + \mathrm{R} \rightleftarrows \mathrm{C}\,, \tag{10.4}$$

and let k_+ and k_- be the association and dissociation rate constants, and K the thermodynamic constant of the reaction ($K = k_+/k_-$). The latter governs the ratio of the concentrations of the various reagents at equilibrium, and depends on the standard enthalpy $\Delta_\mathrm{r}H$ and entropy $\Delta_\mathrm{r}S$ of the reaction:

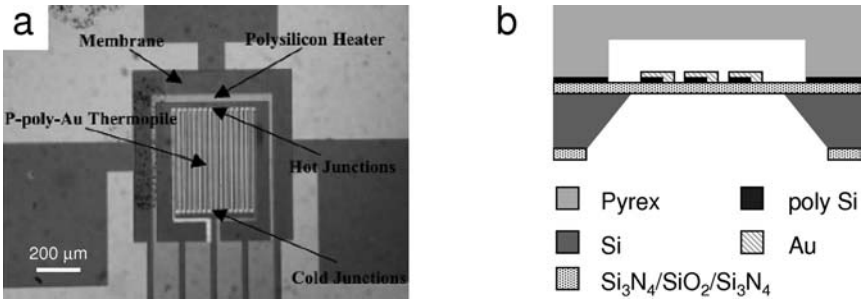

Fig. 10.2 Nanocalorimeter integrated in a microfluidic channel. (**a**) Photographic view from above. (**b**) Cross-section scheme. From [34]

$$K = \frac{[\mathrm{C}]_{\mathrm{eq}}}{[\mathrm{L}]_{\mathrm{eq}}[\mathrm{R}]_{\mathrm{eq}}} = \exp\left(-\frac{\Delta_{\mathrm{r}}H - T\Delta_{\mathrm{r}}S}{RT}\right) . \tag{10.5}$$

The titration technique known as isothermal calorimetry (ITC) consists in measuring the amount of heat Q_i released by the gradual addition of a solution of L to a solution of R. At the ith aliquot, one has $Q_i = n_i \Delta_{\mathrm{r}} H$, where n_i is the amount of complexes to have formed. Since n_i depends on K, the thermodynamic constants $\Delta_{\mathrm{r}}H$ and $\Delta_{\mathrm{r}}S$ can thus be determined, along with the stoichiometry of the reaction, by fitting the curve $Q_i(i)$ [29].

Today, miniaturised calorimeters work with volumes in the range 50–500 nL. They are fabricated around a thin membrane on which the sample is deposited (see Fig. 10.2). This holder, made from polyimide of thickness ~ 10 µm [31] or silicon nitride of thickness ~ 1 µm [32–37], is designed to insulate the solution thermally. The temperature change ΔT caused by the reaction, between 0.1 and 10 mK, is measured using a resistive thermal detector (RTD) [31, 36] or a thermopile, themselves microfabricated at the center of the membrane. More precisely, an RTD (a single conducting wire [31] or a Schottky junction [36]) effectively gives ΔT from the calibration curve of its electrical resistance as a function of temperature, this data being previously obtained by carrying out measurements on a globally thermostated microsystem. The power P dissipated by the reaction, in the range 10–100 nW, is then inferred from equations similar to (10.7) and (10.8). Conversely, for systems based on thermocouples, the type of calibration described above is impossible because the two junctions must be held at different temperatures. Then P is determined directly, after calibrating the setup using a resistive heater, itself deposited on the membrane [32–35], or a chemical reaction with known $\Delta_{\mathrm{r}}H$ [37].

The response time τ_{thermal} of a nanocalorimeter can be as low as 10 ms [32, 36], with miniaturisation playing a major role in improved performance. Indeed, if L is the characteristic size of the device, its total heat capacity C, i.e., including the sample, goes as L^3, while its thermal conductance K goes as L. As a consequence,

$$\tau_{\text{thermal}} \sim \frac{C}{K} \propto L^2 \,. \tag{10.6}$$

It is interesting to estimate this time constant because it provides a way of identifying the effect of size reduction on the temperature variation [35]. For very fast chemical reactions ($\tau_\chi \ll \tau_{\text{thermal}}$, see (10.10) for the definition of τ_χ), the diffusion term can be neglected in the heat equation and, since Q increases in proportion to L^3, this implies

$$\Delta T \sim \frac{Q}{C} \propto 1 \,. \tag{10.7}$$

Consequently, smaller sample volumes do not hamper measurements. However, if heat production is continuous, i.e., for a slow reaction ($\tau_\chi \gg \tau_{\text{thermal}}$), or for example when monitoring the metabolism of a single living cell [32], one has a quasi-steady state regime:

$$\Delta T \sim \frac{P}{K} \propto L^2 \,. \tag{10.8}$$

Miniaturisation thus has some drawbacks too, since in this case it involves the fabrication of ever more sensitive sensors.

Applications of microscale heat transfer to the characterisation of chemical reactions are not only concerned with thermodynamics. For example, the laser T-jump technique is one of the most widely used for measuring ultrafast reaction rates [38, 39]. In this relaxation method [40], the reaction mixture is pushed out of equilibrium by heating with an infrared pulse ($\Delta T \sim 5$ K). Since K depends on T through the Van't Hoff law [the derivative of (10.5)],

$$\frac{\mathrm{d}\ln K}{\mathrm{d}T} = \frac{\Delta_\mathrm{r} H}{RT^2} \,, \tag{10.9}$$

the chemical system evolves into a new composition when its temperature changes. Integrating the kinetic equations for small perturbations then leads to an exponential relaxation of the reagents concentrations [41], with characteristic time

$$\tau_\chi = \frac{1}{k_+\left([\mathrm{L}]_{\text{eq}}^{\text{final}} + [\mathrm{R}]_{\text{eq}}^{\text{final}}\right) + k_-} \,. \tag{10.10}$$

In order to be able to determine the rate constants, the temperature change must be instantaneous for the reactive system ($\tau_{\text{jump}} \ll \tau_\chi$) and the solution must remain isothermal throughout the chemical relaxation ($\tau_{\text{thermal}} \gg \tau_\chi$). The first T-jump apparatus ever built used Joule heating, obtained by discharging a capacitor in the solution. As a consequence, their temporal resolution was limited by the time constant of the electrical setup, which gave $\tau_{\text{jump}} \sim 1$–$10$ μs [42]. With laser heating, it is now the pulse width that specifies the system performance, with $\tau_{\text{jump}} \sim 1$–$10$ ns [43]. Note that, after absorbing the infared radiation ($\lambda_{\text{exc}} \sim 1.5$ μm for H_20), thermalisation between the solvent and the solute molecules takes only 5 ps and so cannot be a limiting factor [43, 44]. With regard to microsystems, no device based on a heating resistor has yet been designed. Since the rise time τ_{jump} is determined

by thermal diffusion, it falls off as L^2. In order to achieve relaxations faster than 10 μs in water, one would therefore have to be able to carry out measurements within submicrometric volumes.

10.2 Microfluidics and Thermometry

We shall now discuss the various thermometric methods used in microfluidics (see Table 10.1). Quite generally, temperature measurement techniques fall into three categories [45, 46]:

- Invasive. The sensor is in direct contact with the medium.
- Semi-invasive. The measurement is carried out at a distance after treating the medium with an ad hoc probe.
- Non-invasive. The measurement is carried out at a distance without pretreatment.

Molecular thermometers, which we shall be discussing below, belong to the second category (see Sects. 10.3–10.7). They were developed to facilitate remote thermal imaging and bypass some of the limitations, e.g., difficulty in implementation, lack of spatial resolution, characterizing totally non-invasive methods such as nuclear magnetic resonance (NMR), Raman spectroscopy, and IR thermography (see Sects. 10.2.2 and 10.7). In addition, these molecular probes have proved essential for the development of new microsystems. Indeed, experimental measurement of the whole temperature field is often needed to make up for the limitations of numerical approaches (excessively complex geometry, advection phenomena, and poorly defined boundary conditions). It is only when heat exchanges throughout the whole device have been understood that a temperature servocontrol can be based on miniaturised, and thus only slightly invasive, electrical sensors (see Sect. 10.2.1).

10.2.1 Electrical Methods

The first advantage in microfabricating temperature and heat flux sensors is to strengthen the integration that characterises the lab-on-a-chip. In addition, miniaturisation can increase the spatial resolution of a measurement, which is proportional to the characteristic size L of the thermometer. The perturbation induced by adding a thermal mass, proportional to L^3, is also smaller, and finally, the response time is improved ($\tau_{\text{thermal}} \sim C/K \propto L^2$). However, macroscopic electrical devices are still used, because many thermometric methods require a calibration curve providing the value of the observable as a function of temperature. The accuracy of the measurement of T at this stage is often what limits the overall accuracy of a technique (~ 0.1 K).

As the principles of electrical measurement techniques will be described in Chap. 18, we shall just give a brief overview of the literature here. Undoubtedly

Table 10.1 Typical characteristics of the main thermometric methods used in microsystems. For the mode, NC = non-contact, C = contact, MP = molecular probe. Regarding the accuracy, it is usually given by the macroscopic probe used for calibration, i.e., $\approx \pm 0.2$ K

Method	Mode	Observable	Relative sensitivity $[10^{-2}\mathrm{K}^{-1}]$	Spatial resolution [μm]	Temporal resolution [s]	Reference
Interferometry	NC	Refractive index	0.01[a]	75	< 0.1	[55]
RTD	C	Resistance	0.3–0.6[a]	30	0.1	[19]
Thermocouple	C	Voltage	0.8[a]	25	5×10^{-8}	[48]
Cholesteric LC	MP	Hue		20	0.01–0.1	[77, 79]
Nematic LC	MP	Optical texture		5	0.1	[77, 80, 81]
Spectrophotometry	MP	Absorbance	0.025			[63]
Fluorimetry	MP	Intensity	2[b]	1	3×10^{-2}	[104]
Raman spectroscopy	NC	Intensity ratio	0.7[b]	1–5	2	[192]
NMR	NC	Chemical shift	0.1	400	1	[201, 203]

[a] Signal varying linearly with T: the relative sensitivity $\mathscr{S}$ is nearly constant over the considered operating range.
[b] Signal varying exponentially in $1/T$. As the relative sensitivity $\mathscr{S}$ is temperature dependent, its value is given here at 40°C.

because they are easy to make, resistive sensors are the most widely used, especially for devices designed for PCR [14]. They exploit the change in electrical resistivity with temperature (~ 3–6×10^{-3} K^{-1}) and take the form of conducting wires made from Ni [47], Pt [19], Au [5], Si [31], ITO [51], etc., with widths of a few tens of microns and thicknesses a few hundred nanometers. The measurement is carried out with a two-point setup, or a four-point setup if one wishes to eliminate contact resistances [19].

Thermopiles, which are sets of thermocouples connected in series, represent the other large family of thermal detectors. They are particularly widely used in nanocalorimetry, and once again can be made from many different pairs of conductors, e.g., Ti/Bi [33], Au/Ni [32], BiSb/Sb [35], polySi/Au [34], chromel/alumel [48]. The possibility of making submicron junctions has led to response times shorter than the microsecond [49, 50].

10.2.2 Non-Spectroscopic Optical Methods

As we have just seen, the integration of micrometric thermometers provides a way of obtaining fast and accurate measurements. However, this data only concerns the immediate vicinity of the sensor. Moreover, a certain number of technological steps have to be added to the chip fabrication process. As a consequence, it may be preferable to develop optical thermometric methods, able to produce, from a distance, a thermal map of an extended field of observation.

The main technique employed for thermal characterisation of electronic devices is infrared thermography, which relies on the light emitted by every heated body [45, 46]. However, the fact that it uses wavelengths in the range 0.7–20 μm greatly restricts its application to the study of microlaboratories. Indeed, for one thing, biochemical reactions take place in water, a solvent that is highly absorbing beyond 1 μm, and for another, resolution is limited to a few microns by diffraction. To our knowledge, apart from several checks of uniform heating [51, 52], only two references are based on infrared thermography [18, 53]. The radiation emitted by the cover of the microdevice is detected by an IR sensor and the temperature of the aqueous mixture deduced from a preliminary calibration procedure. The measurement can only be indirect here because the emissivity properties of the materials making up the chip become poorly defined after a certain number of microfabrication steps.

In order to be able to measure the temperature of the solution of interest directly, one must use some observable that varies monotonically as a function of the temperature. For example, since the refractive index falls off significantly due to thermal expansion ($\mathrm{d}n/\mathrm{d}T \sim -10^{-4}\ \mathrm{K}^{-1}$), many interferometric techniques have been devised to measure variations in the optical path caused by heating. After calibrating the optical signal against a thermocouple, the temperature of the liquid contained in a microchannel was thus determined by backscatter interferometry [54, 55], and Fabry–Perot interferometry [56]. In a similar way, changes in the refractive index have been employed to quantify heating after absorption of laser radiation, either by water during optical trapping experiments [57], or by nanometric metal colloids in the development of a new imaging technique [58]. Finally, one may mention thermal lensing which is used in the detection of non-fluorescent molecules via the photothermal effect. Here a photon is absorbed by the analyte, and then non-radiative relaxation phenomena lead to local heating of the solution and a modification of the refractive index. A diverging lens is thus formed, as revealed by its effect on the focusing of a laser beam [59–61].

10.2.3 Molecular Probe Related Methods

Although relatively powerful, thermal imaging techniques based on refractive index measurements are rather difficult to implement. This no doubt explains why molecular probes are so widely used in on-chip thermography. Indeed, after introducing the thermosensitive species into the device and recording an image by optical microscopy, the temperature at any point is simply calculated with reference to a previously established calibration curve. Developed since the 1960s, thermochromic liquid crystals (LC) were among the first molecular sensors designed for thermal imaging [79]. However, the fact that they are prepared in the form of colloids with diameters above 10 μm makes them difficult to use in microchannels and limits spatial resolution. As a consequence, attention turned to probes that could be directly diluted in the liquid medium whose temperature was to be measured.

Since their absorbance was not sensitive enough, thermochromic probes exploiting transition metal complexes [62, 63] were quickly superseded by fluorescent probes. With a spatial resolution of less than a micron and a response time often less than the millisecond, a whole range of molecules [64–66] and inorganic colloids [67] are now available for use.

In the next three sections, we shall review the various light-emitting probes referenced in the literature (see Figs. 10.3–10.5, 10.7–10.9, 10.11, and 10.12 for examples of molecular structures). Since the chemical reaction (in the broad sense of the term) has been chosen to organise this review [65], we shall describe successively how phase transitions, the kinetics of photophysical phenomena, and thermodynamics can be used to build luminescent molecular thermometers (LMT). In addition, Tables 10.2 and 10.3 provide practical information for selecting the appropriate probe for a given problem. A good thermometer necessarily fulfills a compromise between many relevant properties, e.g., absorption and emission wavelengths, brightness, sensitivity, solubility, and so on.

10.3 Thermosensitive Materials

The physicochemical properties of most materials vary with temperature, and various techniques can thus be devised for thermometry. For instance, analysis of some tracers Brownian motion by video imaging [69, 70] or FCS [71] have been implemented to measure viscosity η and thus retrieve the solvent temperature using tabulated data from the literature. However, to achieve the highest sensitivity one has to favor sharp variations of the observable with T. This can be achieved using phase transitions. In the following, we introduce three molecular systems that have been turned into thermosensitive materials.

10.3.1 Liquid Crystals

The liquid crystal state is characterised by a liquid behaviour in at least one space direction, and by a certain degree of anisotropy. For example, between a solid state at low temperature and an isotropic liquid state at high temperature, rod-shaped molecules can exhibit a nematic phase, in which their center of gravity diffuses freely in three dimensions but their major axis of symmetry follows a unique axis called the director (see Figs. 10.3a and b′). Over a specified temperature range, one thus has an anisotropic arrangement of the substance, reflected on the macroscopic level by remarkable optical properties. Two different types of molecule have been used for thermal mapping of microfluidic reactors: cholesteric LCs [24, 72–75], and nematic LCs [76]. An accuracy of 0.1–0.5 K is generally obtained, with a response time of around 0.1 s.

Table 10.2 Characteristics of several fluorescent kinetic probes used in thermometry. Values are quoted directly from the cited reference, unless otherwise indicated. Concerning the solvent, M = solid matrix, polymer matrix, or other, A = aqueous buffer, and O = organic solvent. Indicated values of λ_{exc} and λ_{em} correspond to experimental conditions used for thermometry. The maxima of the absorption curve and emission curve are given in brackets in the λ_{exc} and λ_{em} columns, respectively. Regarding ϕ and τ_χ, as the kinetic constants depend on the temperature through the Arrhenius law, values are given near room temperature (20–25°C). Concerning the type of measurement, I = intensity, IR = intensity ratio, LT = lifetime. Additional characterisation techniques described in the literature are given in brackets. Concerning the relative sensitivity $\mathscr{S}$, maximal and minimal values are given, unless only one value is indicated, in which case this is calculated in the middle of the operating range. Concerning the reaction, CC = conformational change, ISC = singlet–triplet intersystem crossing, CT = transition via an intermediate with charge transfer, EXC = excimer or exciplex formation

Probe	Solvent	Working range [°C]	λ_{exc} [nm]	λ_{em} [nm]	ϕ [10^{-2}]	E_a [kJ/mol]	τ_χ [s]	Type of measurement	Accuracy [K]	Relative sensitivity [$10^{-2}K^{-1}$]	Type of reaction	Reference
Rhodamine B	A	10–90	500–550	> 565	31[a]			I	0.03–3.5	2.1[a]	CT	[104]
Rhodamine 3B	A	9–77	(558)	(583)	28	25.2	2×10^{-9}	LT			CT	[115]
n-propylamino-NBD	O	20–65	(478)	(545)	39	15	1.2×10^{-8}	LT		1.2[a]	CT	[88]
$Ru(bpy)_3^{2+}$	A	20–60	470	> 550	4.5	42.08	1.5×10^{-5}	I	0.05	1.5–3.9	CT	[124]
$Ru(phen)_3^{2+}$	M (A)	0–50	470	> 570			10^{-6}	LT	0.5	0.1	CT	[122]
$Eu(tta)_3$	M	10–60	345–380	612				I	0.01	5	CT	[133]
$Eu(tta)_3L$[b]	M (O)	0–70	405	615	39	44.8	5.5×10^{-4} [a]	LT		1.4[a]	CT	[132]
DPP	M	0–85	350	(375/494)		15.2		IR (LT)	1	1.5	EXC	[64]
Anthracene-anisidine	O	−150–23	337	395/560	0.87/43	12.13	6×10^{-10}	IR		2.7[a]	EXC	[139]
Acridine yellow	M	−50–50	410	500/580		29.6	0.3	I (IR, LT)	1	2.0–4.5	ISC	[143]
C_70	M	−80–140	470	700	7.1	26	2.7×10^{-2}	IR (LT)		0.05–3.5	ISC	[144]
DNA–RhG[c]	A	15–35	497	524		12.5	2.4×10^{-9}	LT		1.7[a]	CC	[155]

[a] Value either calculated from data given in the cited reference or corresponding to commonly accepted orders of magnitude in the literature.
[b] L is a dipyrazotriazine derivative.
[c] RhG is rhodamine G.

Table 10.3 Characteristics of several thermodynamic probes used in thermometry (luminescence, Raman, or NMR spectroscopy). Values are quoted directly from the cited reference, unless otherwise indicated. All probes operate in water except PMPBA and BTBP, diluted in dodecane and methanol, respectively, Nd^{3+} included in silica, and cystine used in the solid state. Indicated values of λ_{exc} and λ_{em} correspond to experimental conditions used for thermometry. The maxima of the absorption curve and emission curve are given in brackets in the λ_{exc} and λ_{em} columns, respectively. Regarding τ_χ, as the kinetic constants depend on the temperature through the Arrhenius law, values are given near room temperature (20–25°C). Concerning the type of measurement, I = intensity, IR = intensity ratio, LT = lifetime. The relative sensitivity $\mathscr{S}$ is determined in the middle of the operating range. Concerning the reaction, CC = conformational change, COM = complexation, P = protonation, T = thermalisation between excited states, EXC = excimer or exciplex formation, HB = hydrogen bonding

Probe	Working range [°C]	λ_{exc} [nm]	λ_{em} [nm]	τ_ϕ [s]	$\Delta_r H$ [kJ/mol]	$\Delta_r S$ [J/mol K]	τ_χ [s]	Type of measurement	Accuracy [K]	Relative sensitivity [$10^{-2}K^{-1}$]	Type of reaction	Reference
Poly(DBD-AE-co-NIPAM)	29–37	458	(566/533)	10^{-7}–10^{-10} [a]	6.3[a]	20.6[a]	< 100[a]	I		21.5[a]	CC	[83, 89]
DMABN⊂ βCD	10–35	295	350/530	10^{-7}–10^{-10} [a]	−103	−74.7		IR	2.5	6[a]	COM	[161, 162]
BN⊂[βCD•ROH]	20–50	308	(480–550)	2×10^{-2} [a]				LT	0.3	18.5	COM	[158]
N-βCD	0–80	295	376[a]	10^{-7}–10^{-10} [a]	−29.8	−103.2		I		2.4[a]	COM	[160]
Molecular beacon	7–37	495	525/625	10^{-7}–10^{-10} [a]	−169	−582	5×10^{-5}	IR		7	CC	[148]
BCECF	20–80	(482/503)	(520/528)	10^{-7}–10^{-10} [a]				I		0.95	P	[163, 169]
PYMPON	3–63	341	(450/550)	10^{-7}–10^{-10} [a]	−23		10^{-8}	IR		2	P	[148]
Tripodal polyamine	10–60		324/399	10^{-8}–10^{-9}	−45	−128	7.7×10^{-9}	IR		4.5	EXC	[138]
PMPBA	20–100	344	396/448	10^{-7}–10^{-10} [a]	13.1			IR	1	2[a]	EXC	[140]
Nd^{3+}	−50–500	802	830/905		12.32[c]			IR	3	1.68	T	[179]
BTBP	15–70	480	536/550	10^{-7}–10^{-10} [a]	5.68[c]		10^{-12}–10^{-10} [a]	IR	2	0.4[a]	T	[176]
Cystine[b]	20–110	514	501/527	10^{-15} [a]	11.96[c]		10^{-12}–10^{-10} [a]	IR	2		T	[191]
H_20[b]	20–65	532	645/654	10^{-15} [a]	11[a]		10^{-12} [a]	IR	1	0.7	HB	[192, 193]
H_20	20–85	RMN^{1}H		10^{-3} [a]	11[a]		10^{-12} [a]	'IR'	0.2	0.1	HB	[202, 203]

[a] Value either calculated from data given in the cited reference or corresponding to commonly accepted orders of magnitude in the literature.
[b] Measurements performed by Raman spectroscopy.
[c] These values correspond to ΔE, the energy gap separating the two emitting sublevels.

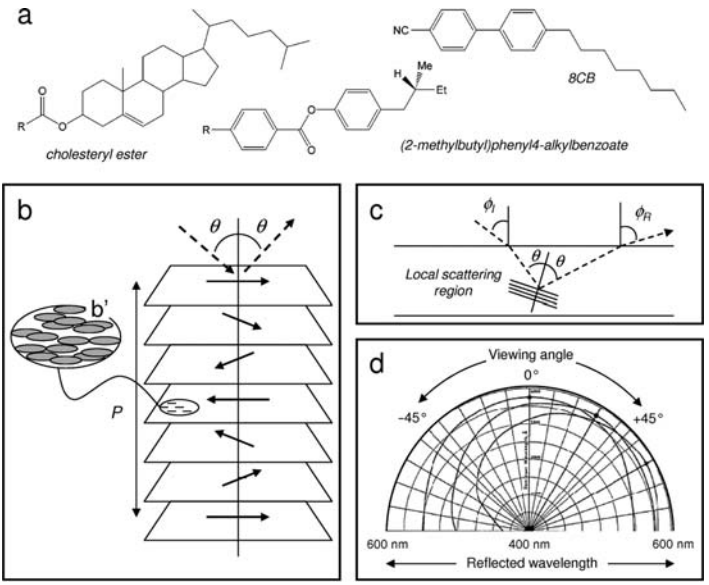

Fig. 10.3 (**a**) Molecular structures of compounds able to form either a cholesteric phase (esters of cholesterol, (2-methylbutyl)phenyl4-alkylbenzoates), or a nematic phase (8CB). (**b**) Arrangement of molecules in a cholesteric liquid crystal film. (**b′**) Local nematic order. (**c**) Bragg reflection and diffraction on a multidomain liquid crystal. For a given illumination angle ϕ_I and viewing angle ϕ_R, only the wavelength $\lambda = \bar{n}P\cos\left[(1/2)\arcsin(\sin\phi_I/\bar{n}) + (1/2)\arcsin(\sin\phi_R/\bar{n})\right]$ will be reflected. (**d**) Colour of a cholesteric film as a function of the viewing angle for three fixed illumination angles. Taken from [77]

Cholesteric Liquid Crystals

The presence of an asymmetric center in a rod-shaped molecule can lead to gradual rotation, along a helix of pitch P, of the director of the nematic order. The molecules thus remain parallel to one another only within a given plane (see Figs. 10.3a and b). Optically, the thin films obtained are characterised by considerable rotating power, polarisation properties, and iridescence [68,77]. The latter effect arises because, to a first approximation, the liquid crystal can be viewed as a stack of nematic layers with periodically modulated refractive index. Only the wavelength λ given by the first order Bragg law is then back-scattered, viz., $\lambda = \bar{n}P\cos\theta$, where $\bar{n}$ is the average index of the film and θ the angle between the incident beam and the helix axis. Moreover, the wavelength of maximum reflectivity shifts from the red to the violet when the temperature increases. Indeed, the main consequence of more pronounced Brownian motions is an increase in the angle formed by the directors belonging to successive nematic layers, and hence a reduction in P.

In practice, thermochromic compounds have the form of suspensions of polymer capsules containing the liquid crystal [78]. The spatial resolution of the measurement is then given by the diameter of the particles, in the range 10–150 μm. Furthermore, depending on the mixture of cholesteric compounds used in the formulation, the shift from red to violet can occur over a temperature range going from

a few tenths of a degree to about twenty degrees, whence the measurement sensitivity can be adjusted. The experimental setup required for thermometry comprises a white light source, a light sensor, and the device to be tested filled with liquid crystal (a black background is recommended to reduce unwanted reflections) [79]. The temperature is then inferred from the probe hue [24,74,79] or its wavelength of maximum reflectivity [73]. Since these parameters are highly sensitive to the lighting and viewing angles (see Figs. 10.3c and d), in situ calibration is always needed [77,79].

Nematic Liquid Crystals

The molecular order present in nematic films makes them birefringent. They thus appear bright when imaged between crossed polarisers. This property can be used to observe the transition to the isotropic phase, which is completely dark. More precisely, on a device coated with a nematic film, the line of high black/white contrast corresponds to the isotherm $T = T_{\mathrm{n}\to\mathrm{i}}$, where $T_{\mathrm{n}\to\mathrm{i}}$ is the temperature at which the phase change occurs [76,80]. This technique involves no artifacts due to the type of optical setup and requires no calibration, since $T_{\mathrm{n}\to\mathrm{i}}$ is a thermodynamic property of the chosen liquid crystal. Moreover, by artificially varying the temperature of the surrounding medium by δT, one can image what would otherwise be the isotherm $T = T_{\mathrm{n}\to\mathrm{i}} - \delta T$. A single nematic compound, such as 8CB (see Fig. 10.3a), can thus be used to obtain a full thermal map [81,82].

10.3.2 Polymers

Some polymers such as poly(N-isopropylacrylamide) (or poly(NIPAM), shown in Fig. 10.4a) change from a statistical coil state to a globule state when the temperature goes above a certain value Θ_{LCST}, called the lowest critical solution temperature [83]. This phenomenon can be viewed as a consequence of a shift in some equilibrium, where monomers change from a hydrophilic form to a hydrophobic form. There is then a significant change in the interactions of the polymer chain with water molecules [84], the latter being expelled from the coil. Consequently, the density of these nano-objects increases by a factor of about 100 (see Fig. 10.4b) [85–87].

Since the form and position of the coil–globule transition are perfectly specified for each polymer system, there have been applications to molecular thermometry. These are all identical in principle, differing only through the process used to observe the compaction of the chain. For example, when copolymerised with thermosensitive monomers, some benzofuran derivatives have their fluorescence quantum yield increased by a factor of more than 10 by lowering the temperature below the value Θ_{LCST}. Indeed, these probes are particularly sensitive to the polarity of their environment [88], which decreases significantly with the formation of the globule [89,90]. In a similar way, pyrene residues can form excimers and thus be

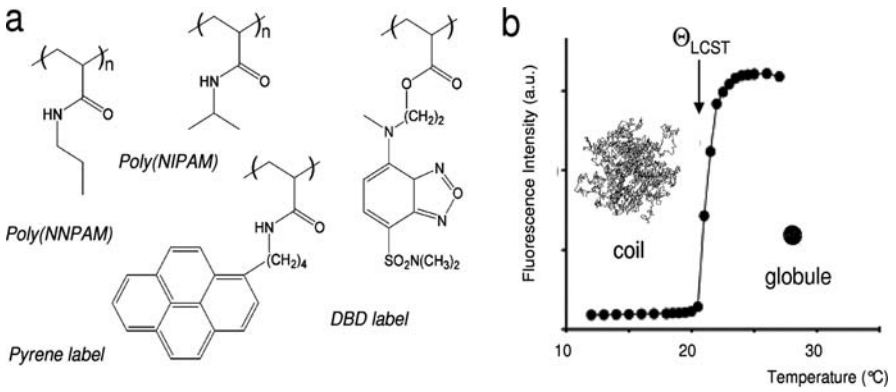

Fig. 10.4 (**a**) Molecular structures of monomers that can be used to synthesise a fluorescent thermosensitive polymer. The signal is emitted by a probe, either pyrene or DBD, diluted in a matrix which exhibits an LCST, either poly(NIPAM) or poly(NNPAM). (**b**) Temperature dependence of the fluorescence intensity for poly(NNPAM) labelled with DBD (0.01% m/v in water, $\lambda_{exc} = 444$ nm, $\lambda_{em} \sim 530$–560 nm). Taken from [89]

used to reveal conformational changes induced by temperature modifications (see Sect. 10.4.3) [91, 92].

As with fluorophores, the optical response of nanoparticles can also be modified by polymer phase transitions [93]. In poly(NIPAM) gels, compaction of the lattice causes a change in the refractive index of the environment and brings together the colloids diluted within it. This in turn leads to variations in the intensity and position of the photoluminescence band for quantum dots [94] and the plasmon resonance band for gold particles [95]. In an even more sophisticated system, a large gold colloid was coated with smaller quantum dots with the help of an intermediate layer of poly(ethylene glycol). Since the volume of this polymer changes with temperature, the exciton–plasmon coupling is also modulated, which yields a photoluminescence signal of different intensity [96, 97].

10.3.3 Phospholipid Membranes

The membrane of a liposome is a 2D molecular assembly with a structure that is extremely sensitive to temperature. Hence, as in 3D systems, phase transitions have been observed. The transition from a gel state to a liquid state can be exploited in thermometry, like the coil–globule transition described above. One need only use an appropriate reporter probe. Since the occurrence of the liquid phase leads for the lipid bilayer to increased fluidity and greater permeability to water, two fluorescent molecules sensitive to the environment have been selected, namely, laurdan and a NBD conjugate (see Fig. 10.5a) [98, 99]. Following a rise in temperature, the first does indeed exhibit a drop in quantum yield and a bathochromic shift (i.e., to

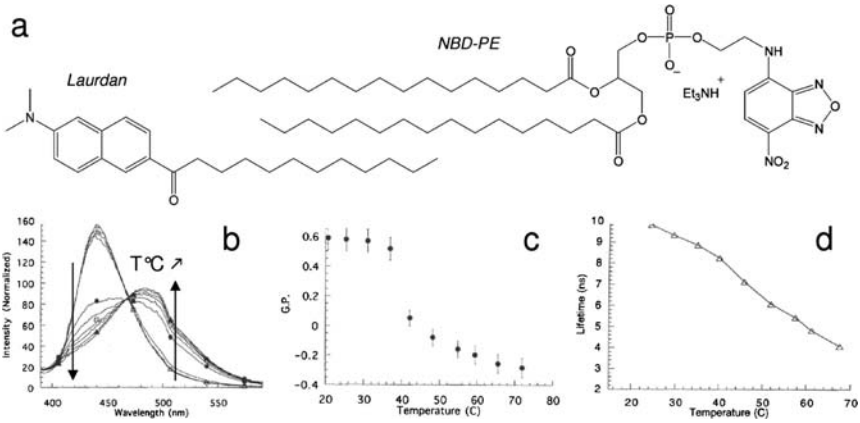

Fig. 10.5 (**a**) Molecular structures of two fluorescent membrane probes used for thermometry, laurdan and NBD-PE. (**b–d**) Temperature dependence of the fluorescence signal of a suspension of dipalmitoylphosphatidylcholine vesicles labelled by these two probes. (**b**) Fluorescence spectra of laurdan when T varies between 21 and 72°C. (**c**) Corresponding generalised polarisation GP $= (I_{440} - I_{490})/(I_{440} + I_{490})$, where I_{440} and I_{490} are the fluorescence intensities measured at 440 and 490 nm, respectively ($\lambda_{\text{exc}} = 365$ nm). (**d**) Slowest component of the NBD-PE fluorescence lifetime ($\lambda_{\text{exc}} = 475$ nm). Taken from [99]

longer wavelengths) in emission, reflecting an increase in polarity (see Fig. 10.5b). A ratiometric measurement at wavelengths characteristic of both the gel and liquid phases can then be used to determine the temperature independently of concentration fluctuations (see Fig. 10.5c). The derivative of NBD for its part only exhibits a reduction in quantum yield, reflecting a faster non-radiative relaxation in the liquid phase, certainly due to a reduction in the local viscosity. Concentration-independent thermometry can then be carried out by measuring the fluorescence lifetime (see Fig. 10.5d).

10.4 Kinetic Fluorescent Probes

In this section we discuss fluorophores whose emission after photon absorption is controlled by the kinetics of the de-excitation processes. In other words, this means that the radiative lifetime of the probe is then comparable with the characteristic decay time τ_χ of another relaxation channel. In addition, as we are focusing on molecules used in thermometry, the relevant rate constants will be strongly temperature dependent.

More precisely, absorption of a photon takes the luminescent probe from a singlet ground state S to an excited singlet state S* (see Fig. 10.6) [100, 101]. Several de-excitation pathways are then available for the molecule to return to its initial state. They involve a certain number of possible intermediate states, e.g., triplet state T*, excited product U*, as well as various photophysical mechanisms, e.g., fluorescence, phosphorescence, internal conversion, intersystem crossing. Each of

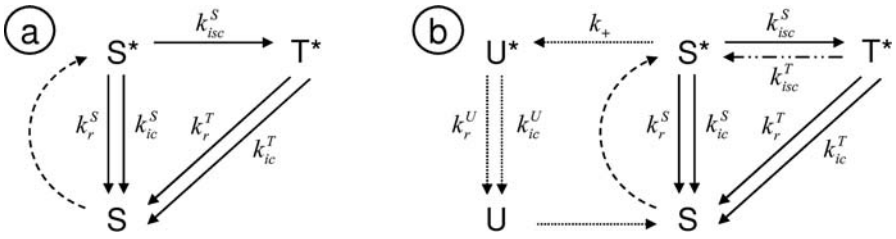

Fig. 10.6 After absorption of a photon (*dashed arrows*), a photochemically stable fluorescent probe can de-excite along various photophysical pathways. (**a**) Simple case where, starting from the excited singlet state S^*, there are only three possibilities: radiative relaxation *r* and internal conversion *ic* to the ground state S, and intersystem crossing *isc* to the triplet state T^* (*continuous arrows*). Once formed, the latter can in its turn relax radiatively and non-radiatively to S (*continuous arrows*). (**b**) More complex case where there is also a possibility of reverse intersystem crossing from the triplet state to the singlet state (*dot-dashed arrow*), and formation of an excited product U^* (*dotted arrows*). Like T^*, U^* will relax to the ground state by its own de-excitation channels (*dotted arrows*)

the latter transitions is characterised by a kinetic constant *k*, and one needs to know the full set of their values to be able to determine the preferred relaxation channel the molecule will follow.

The radiative lifetime of the singlet state lies in the range 1–100 ns, while that of the triplet state is more like one second because the associated transition to the ground state is spin-forbidden. It is also interesting to note that the rate constants k_r associated with the radiative processes generally depend only weakly on the temperature. For their part, internal conversion and intersystem crossing have characteristic times in the ranges 0.01–1 ns and 0.1–10 ns, respectively. Like the formation of excited products U^*, these processes are thermally activated, i.e., by collisions or vibrational excitations. As a consequence, their kinetic constants obey an Arrhenius law of the form

$$k = A \exp\left(-\frac{E_a}{RT}\right) , \tag{10.11}$$

where *A* is a constant and E_a the activation energy of the relevant mechanism.

As we shall see shortly, the two main observables used in luminescence spectroscopy are the quantum yield of emission ϕ and the radiative lifetime τ_ϕ. The first is the number of photons reemitted relative to the number of photons absorbed. The second is the time constant specifying the exponential decay of emission following a pulsed excitation. These two parameters are closely related, because they derive from the same kinetic equation, viz., in the context of fluorescence and for the simple case described in Fig. 10.6a,

$$\frac{\mathrm{d}[S^*]}{\mathrm{d}t} = i_\mathrm{A} - \left(k_\mathrm{r}^\mathrm{S} + k_\mathrm{ic}^\mathrm{S} + k_\mathrm{isc}^\mathrm{S}\right)[S^*]\,, \tag{10.12}$$

where i_A is the concentration of photons absorbed per unit time [101]. Depending on whether one is working in transient or steady-state irradiation conditions, integration leads respectively to

$$\tau_\phi = \frac{1}{k_\mathrm{r}^\mathrm{S} + k_\mathrm{ic}^\mathrm{S} + k_\mathrm{isc}^\mathrm{S}} \tag{10.13}$$

or

$$\phi = \frac{k_\mathrm{r}^\mathrm{S}[S^*]}{i_\mathrm{A}} = \frac{k_\mathrm{r}^\mathrm{S}}{k_\mathrm{r}^\mathrm{S} + k_\mathrm{ic}^\mathrm{S} + k_\mathrm{isc}^\mathrm{S}}\,. \tag{10.14}$$

One can thus see how the quantum yield and luminescence lifetime usually fall off in concert when the rate of some other relaxation channel increases.

10.4.1 Intramolecular Charge Transfer in Organic Molecules

Rhodamine B [102–106] is commonly used for microscale thermal imaging. Like certain molecules of the same family, such as rhodamine 3G [107], its emission quantum yield actually varies significantly with the temperature ($\sim 2 \times 10^{-2}\,\mathrm{K}^{-1}$, see Fig. 10.7). Although its molecular structure was sometimes taken to suggest that

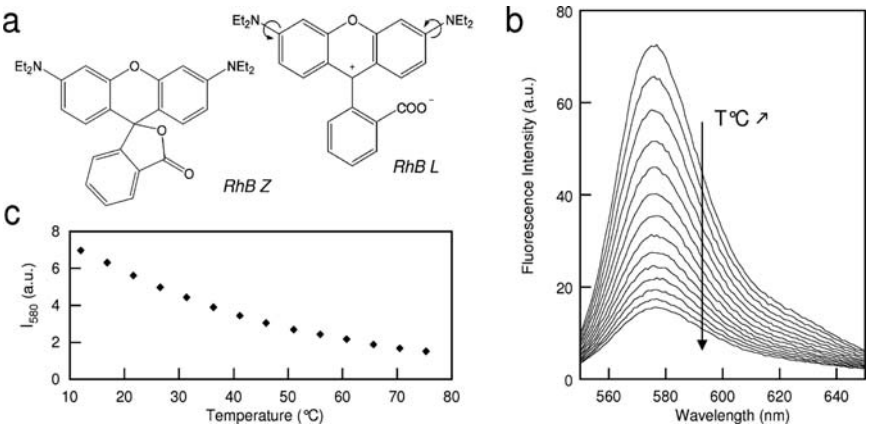

Fig. 10.7 (**a**) Molecular structures of the two isomers of rhodamine B present in aqueous solution at neutral pH: the lactone form RhB L and the zwitterion form RhB Z. *Arrows* indicate rotations likely to be involved in the formation of a TICT state. (**b**) and (**c**) For the same compound and for a temperature varying between 12 and 75°C, temperature dependence of the fluorescence spectrum and the fluorescence intensity at 580 nm ($\lambda_\mathrm{exc} = 541$ nm). Solution at 1 mM in 40 mM $\mathrm{Li(CH_3)_2AsO_2}$ buffer, pH = 7.2 at 25°C

a shift in equilibrium between the fluorescent zwitterion form RhB Z and the non-fluorescent lactone form RhB L might be the origin of this phenomenon, this is not actually the case. Over the relevant temperature range, the relative concentration of each isomer does not vary by more than 10% [108–110]. In the same way, it has been shown that the acid–base properties of the zwitterion form are not relevant here to account for the variation of ϕ with T (its pK_a is 3.1) [111].

The temperature sensitivity of fluorescence emission can in fact be explained by the existence of a photoinduced reaction leading to the formation of a highly polar excited conformer called a TICT (twisted intermolecular charge transfer) state [112]. More exactly, the excitation causes electrons to go from one molecular orbital to another, whence the charge distribution in S^* is not the same as the one in S. The result is then a reorganisation of the solvent molecules and a slight conformational change in the fluorophore. This sequence of events further leads the excited molecule S^* to a state U^* of lower energy, with increased charge transfer and distortion of the geometry (in such a way as to stabilise the charge separation). In the case of rhodamine B, it is the xanthene group that serves as acceptor and the diethylamino groups which donate electrons and rotate during isomerisation [113–115]. Since the latter rearrangement is thermally activated, the rate of formation of the TICT state increases with temperature. The final result is a reduction of the quantum yield since U^* is non-fluorescent.

Conversely, note that the electronic and structural rearrangements characterising probes exhibiting a TICT state also make them particularly sensitive to certain features of their environment, e.g., polarity, viscosity, and so on. This observation has been put to use to make molecular thermometers relying on fluorophores such as NBD [88, 89, 99] or DMABN [162] (see Sects. 10.3.3 and 10.5.3).

10.4.2 Charge Transfer in Organometallic Complexes

As for the organic dyes discussed above, the temperature dependence of the fluorescence quantum yield of many organometallic complexes can be explained by the formation, after charge transfer, of a non-radiative excited state.

Transition Metals

Once complexed by α-diimine ligands, the metal ions of the platinum column (in the periodic table) form luminescent compounds that have remarkable photophysical and photochemical properties [116]. $Ru(bpy)_3^{2+}$ is without doubt the most widely studied of all (see Fig. 10.8a), and it has served as a bridgehead in the development of a whole series of molecular thermometers. There are three main types of electronic transition in this complex. They occur between the ground state and the π–π^*, d–d, or MLCT (metal to ligand charge transfer) states, depending on whether

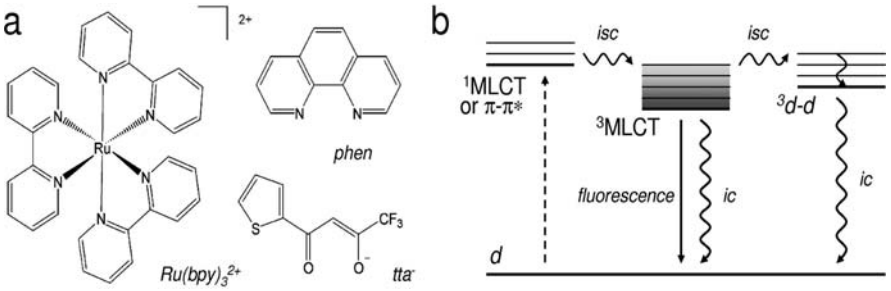

Fig. 10.8 (**a**) Molecular structures of the $Ru(bpy)_3{}^{2+}$ organometallic complex and of the phen and tta^- ligands. (**b**) Diagram showing the different excited states involved in the de-excitation cycle of $Ru(bpy)_3{}^{2+}$

the electron moves respectively from an orbital of the ligand to another orbital of the ligand, from a *d* orbital of the metal to another *d* orbital, or from an orbital of the metal to an orbital of the ligand. As can be seen from the diagram in Fig. 10.8b, after excitation of the π–π^* or MLCT bands, intersystem crossing quickly leads to ^{3}MLCT, the excited states the lower in energy. The latter can relax radiatively, but fluorescence is not the only possible decay mode. Indeed, one can also thermally populate some *d*–*d* states from which subsequent de-excitation to the ground state happens non-radiatively, the relevant transitions being forbidden. Since the access rate to these *d*–*d* states is then proportional to the occupation of the higher vibrational modes of ^{3}MLCT, it increases as an exponential of $-1/T$, typical of the Boltzmann statistics [117–119]. At high temperatures, there is therefore a tendency to favour this reaction at the expense of emission.

Ruthenium complexes like $Ru(bpy)_3{}^{2+}$ and $Ru(phen)_3{}^{2+}$, and also platinum complexes like the Pt-OctaethylPorphyrin [120], have been used to make temperature sensors based either on an intensity measurement [121] or a determination of the luminescence lifetime (decays being in the µs range, experimental implementation is a little easier than when working with organic dyes) [121, 122]. Additionally, these two observables have been used for microscale thermography, after either one- [122–124] or two-photon [125] excitation.

Rare Earths

The photophysical mechanisms making it possible to employ rare earth complexes as molecular thermometers are very similar to those described above [126–129]. $Eu(tta)_3$ is without doubt the most widely studied of the series (see Fig. 10.8a), and we shall use it as an illustration, bearing in mind that other probes have been devised by complexing europium or terbium with β-diketone ligands or cryptands [130].

Since f–f transitions are forbidden, the lanthanides have a very low molar extinction coefficient and they can only be excited via transitions associated with either the ligand or some charge transfer states. Non-radiative relaxation then leads to the excited state of the rare earth ion, the lower in energy, viz., $^5D_0$ in Eu^{3+}. Thereafter, radiative and non-radiative relaxations to the ground state ^7F occur very slowly ($\tau_\phi \sim$ ms), and a second de-excitation channel can be set up. It involves thermal population of some LMCT (ligand to metal charge transfer) states and non-radiative relaxations.

Uses of these complexes in thermometry include fabrication of temperature sensors [131, 132], thermography in microsystems [133, 134], and even measurement of the heat given off by a living cell [135].

10.4.3 Excimer Formation

The collision of a fluorophore in the ground state with another fluorophore in an excited state can lead to the formation of an excited complex. The latter is called an excimer if the two molecules are the same, and an exciplex if they are different. This association phenomenon is diffusion-controlled and can therefore only take place if the two fluorophores are close enough, or if the lifetime of the excited state

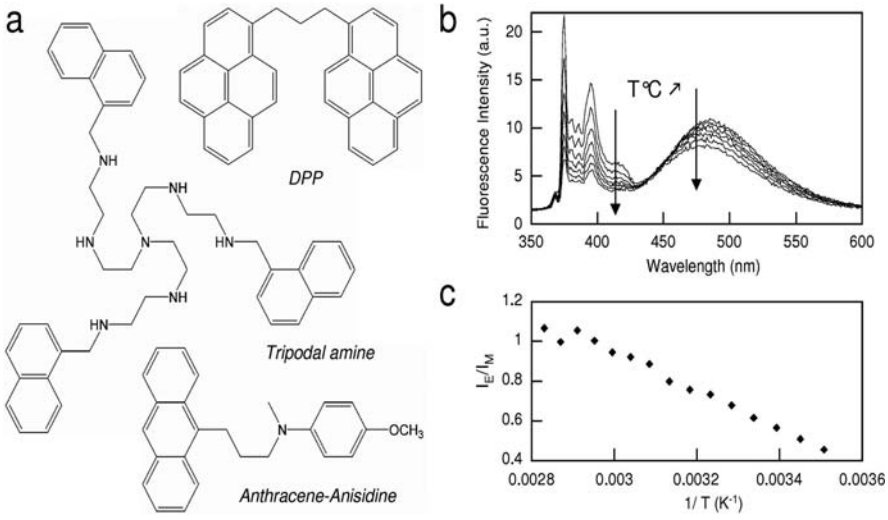

Fig. 10.9 (**a**) Molecular structures of three compounds able to form, after excitation, either intramolecular excimers or intramolecular exciplexes. (**b**) and (**c**) Temperature dependence of the fluorescence spectrum and ratio of fluorescence intensities at $\lambda_{em} = 500$ nm (associated with excimer E) and at $\lambda_{em} = 375$ nm (associated with monomer M) for DPP (solution in dodecane, $\lambda_{exc} = 333$ nm, temperature varying between 12 and 80°C)

is long enough, since a sufficient number of collisions must occur during the fluorescence lifetime. One thus works at high concentrations [23, 136] or by joining several probes together covalently (see Fig. 10.9) [64, 92, 137–139]. Furthermore, given the reaction mechanism involved here, the viscosity and temperature will of course be among the more important parameters controlling this association process [101].

A certain number of unimolecular temperature probes have been developed from aromatic hydrocarbons such as naphthalene or pyrene, these planar structures being prone to form excimers. The operating principle of thermometers obtained in this way can still be described by the diagram in Fig. 10.6b. After absorption, one of the monomeric moieties transits to the excited state S^*. Relaxation then occurs either by fluorescence emission, or by formation of the excimer U^*, which in its turn will relax both radiatively and non-radiatively. Since the excimer emits at a shorter wavelength than the monomer, a two-colour measurement of the population ratio can be made, hence yielding a temperature determination independent of the probe concentration (see Sect. 10.6.2) [23,64,137,138,140]. Once again, note that analysis of the fluorescence decay can also be used for thermometry [64, 138, 141]

Finally, it should be mentioned that excimers do not always work in the kinetic mode (see Tables 10.2 and 10.3). In the above, we assumed that the concentration of the product U^* was controlled by its rate of formation, i.e., we assume that the reverse reaction in which the excimer transforms to an excited monomer; could be neglected. This is only in fact justified at 'low temperatures'. Indeed, in all excimer thermometers there is a 'high temperature' regime where the ratio of concentrations in S^* and U^* is no longer controlled by the kinetics, but by thermodynamics. More precisely, in this latter case, equilibrium between the two excited states can be reached before fluorescence emission has occurred (see Sect. 10.5) [64, 66].

10.4.4 Delayed Fluorescence

Although well known for quite a long time now, delayed fluorescence has not received much attention, mainly because of low intensity levels [101, 142]. While the standard fluorescence mechanism involves radiative decay of a freshly produced excited state, the mechanism here is considerably more complicated. Indeed, emission is now controlled by a second kinetic regime involving the thermal population of S^* from the higher vibrational levels of T^* (see Fig. 10.6b). If the intensity of this phenomenon is not to be negligible, the singlet and triplet states must be close enough in energy and de-excitation by fluorescence and phosphorescence must not be too fast compared with intersystem crossing.

Several temperature sensors have been made by dissolving compounds like acridine orange [143] or the fullerene C_{70} [144] in a solid matrix. They exploit the increase of k_{isc}^{T}, and hence the total fluorescence intensity, with temperature.

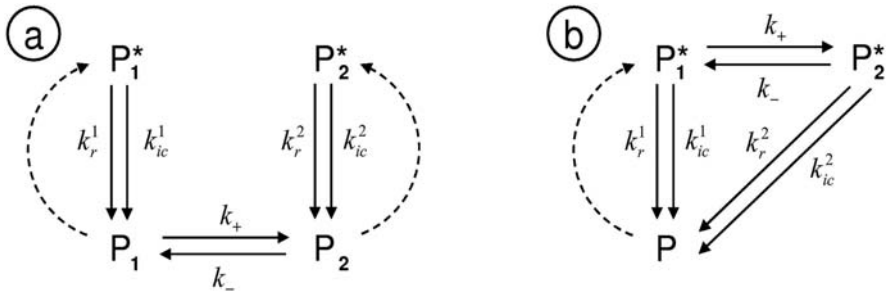

Fig. 10.10 In the simple context of a two-level mechanism, photophysical diagram describing two major types of thermodynamic temperature probe, (**a**) Luminescence is used to measure the extent of an isomerisation reaction involving two ground states. (**b**) Emission properties reflect a thermalisation process occurring between two different excited states or two different sublevels of the same excited state

10.5 Thermodynamic Fluorescent Probes

Thermodynamic probes rely on a temperature-induced shift in equilibrium. Since the species involved are each characterised by their own luminescence properties, the composition of the mixture, whence also the extent of the reaction and the temperature, can be deduced using ad hoc spectroscopic methods.

In this section, we shall discuss two different types of reaction, depending on whether they involve different species in the ground state (Sects. 10.5.1–10.5.5) or different excited states of the same molecule (Sect. 10.5.6). In addition, the discussion below is not restricted to luminescence, and a certain number of general principles can be used to implement other kinds of thermodynamic thermometry (see Sect. 10.7 for measurements involving Raman and NMR spectroscopy).

The temperature determination must only be made once the equilibrium state has been reached. Moreover, it must be faster than the relaxation time of the chemistry in order to obtain a snapshot of the various populations. Regarding the first type of probe, made up of species reacting in the ground state, a spectroscopic technique faster than the reaction is used, i.e., $\tau_\phi \ll \tau_\chi$. If for the purposes of simplicity we consider the case where the excitation does not affect reactivity, the probability of transition from one excited state to the other is then very small, and the photophysical diagram looks like the one in Fig. 10.10a. Regarding the second type of probe, whose photophysical diagram is given in Fig. 10.10b, it has already been discussed in the context of excimers (see Sect. 10.4.3). If the thermalisation between the various excited states is to have time to get established, the decay must now be slower than the relaxation time associated with this transformation, i.e., $\tau_\phi \gg \tau_\chi$.

10.5.1 Isomerisation Between Species in Their Ground State. General Features

As discussed in Sect. 10.1.4, any chemical equilibrium obeys Van't Hoff's law. For a fluorescent probe P that can exist in two isomeric forms P_1 and P_2,

$$P_1 \rightleftarrows P_2 \,, \tag{10.15}$$

equation (10.5) takes the form

$$\frac{d\ln\left([P_2]_{eq}/[P_1]_{eq}\right)}{dT} = \frac{\Delta_r H}{RT^2} \,. \tag{10.16}$$

A change in temperature thus alters the ratio $\mathscr{R} = [P_2]_{eq}/[P_1]_{eq}$ between the populations of the two isomers, and conversely, once $\Delta_r H$ has been determined in preliminary experiments, measurement of variations in $\mathscr{R}$ leads to a determination of variations in T. Most of the thermometric methods presented below are based on this idea. However, in many cases only a signal averaged over the two populations is collected and the temperature measurement thus becomes a little more involved (see Sect. 10.6) [148].

Thermodynamic thermometers, either luminescent or otherwise (see Table 10.3 and Sect. 10.7 for NMR and Raman probes), can be characterised by three main parameters:

- The relative sensitivity $\mathscr{S}$, an expression for which is derived directly from the Van't Hoff law:
$$\mathscr{S} = \frac{1}{\mathscr{R}}\frac{d\mathscr{R}}{dT} = \frac{\Delta_r H}{RT^2} \,. \tag{10.17}$$
- The response time, which in fact is the relaxation time of the chemical reaction $\tau_\chi = 1/(k_+ + k_-)$.
- The population inversion temperature $T_{1/2}$, at which $[P_1] = [P_2]$:
$$T_{1/2} = \frac{\Delta_r H}{\Delta_r S} \,. \tag{10.18}$$

The latter parameter corresponds to the temperature around which the method will give the best results. Indeed, while $\mathscr{S}$ is in theory independent of the concentrations of P_1 and P_2, problems of signal-to-noise ratio and instrument sensitivity make it relatively difficult to measure $\mathscr{R}$ when one of the two species is much less present than the other, i.e., $T \gg T_{1/2}$ or $T \ll T_{1/2}$. When designing a molecular thermometer, it is thus important to maximise $\Delta_r H$ in order to gain sensitivity, while maintaining a value of $\Delta_r S$ which allows one to work around the target temperature. This will necessarily involve using the chemistry of weak bonds, i.e., protonation, complexation, hydrogen bonds, when one intends to carry out measurements in the range 0–100°C. Furthermore, the use of non-covalent chemistry has the advantage of relatively short response times.

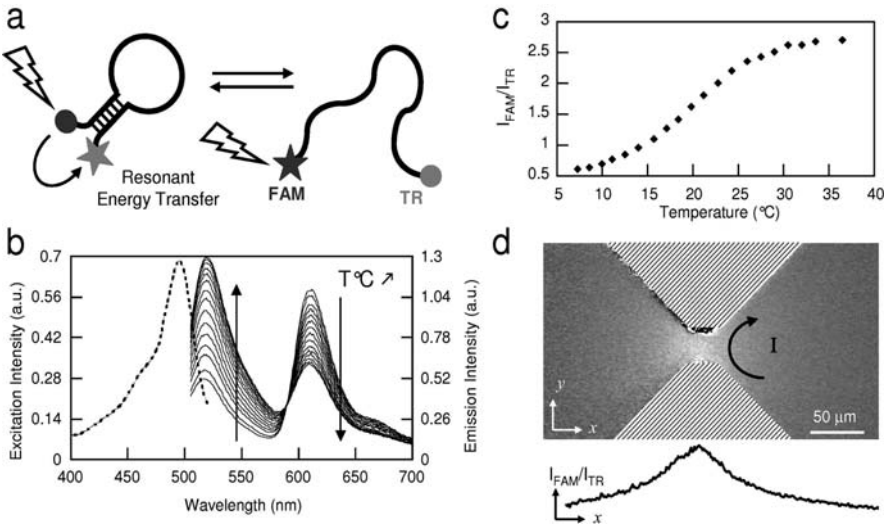

Fig. 10.11 Principle of a thermometer based on a molecular beacon. (**a**) After exciting the fluorescein (FAM), this same fluorophore directly reemits in the open form (dominant at high temperature), whereas it is the Texas Red (TR) which, after resonant energy transfer, emits in the closed form (dominant at low temperature). (**b**) Excitation spectrum of FAM at 25°C and emission spectrum of the beacon as the temperature varies over the range 7.5–36.5°C. Solution at 50 nM in 100 mM NaCl 5 mM NaOH 10 mM HEPES buffer, pH = 7.5 at 25°C, $\lambda_{exc} = 495$ nm. (**c**) Temperature dependence of the ratio between the fluorescence emission intensity at 525 nm (FAM) and the one at 625 nm (TR). (**d**) Thermal map of a resistive element obtained by ratiometric imaging using a dual-view microscope. The current *I* heats, by the Joule effect, a small square of semi-transparent semiconductor placed between two metal contacts (*shaded* on the diagram)

10.5.2 Folding of Nucleic Acid Structures

A molecular beacon [145, 146] is a strand of DNA containing a few dozen bases, labelled at one end by a fluorophore and at the other by a quencher. Its sequence is chosen in such a way that the two ends can hybridise, whence two structures coexist at equilibrium: a fluorescent open form and a non-fluorescent closed form, the shortening of the distance between the fluorophore and the quencher causing extinction of the emission. The fact that the molecular thermometer thereby formed is a simple oligonucleotide with a double label allows a great deal of flexibility in the design. Indeed, these molecules are widely available today by customised synthesis, while the choice of fluorophore enables one to select the excitation and emission wavelengths, and the choice of sequence allows one to select $T_{1/2}$ and $\mathscr{S}$ by adjusting $\Delta_r H$ and $\Delta_r S$. Furthermore, replacing the quencher by a second fluorophore absorbing at the emission wavelength of the first, it is also possible to make thermometers in which the closed form emits after fluorescence resonance energy transfer (FRET) [147,148]. A ratiometric measurement at two wavelengths can then be set up (Figs. 10.11a–c).

The temperature $T_{1/2}$ around which the derivative of the fluorescence signal goes through a maximum is a thermodynamic characteristic of the chosen oligonucleotide. As a consequence, molecular beacons were first used to measure the difference between the temperature given by a macroscopic thermometer stuck onto the microdevices and the true temperature within the corresponding microfluidic chambers [149, 150]. Today more sophisticated applications are envisaged, e.g., in thermal mapping (Fig. 10.11d) [148]. Finally, it is worth noting that, according to reports in the literature, these hairpin-like nucleic acid structures may also be used as temperature sensors by some living organisms [151].

Even though they do not display the flexibility characteristic of oligonucleotide folding, other structural transitions involving biological objects have also been exploited to make molecular thermometers, e.g., transition from the B to the Z form in DNA [152], formation of a DNA/protein complex [153], α-helix to 3_{10} helix transition for a peptide [154]. In addition, it has been shown that, for a rhodamine bound to the end of an oligonucleotide, the acceleration of conformational reorientations with temperature favours quenching by nucleic bases [155]. A thermometer based on fluorescence lifetime measurements was thereby built. However, note that the latter example lies outside the thermodynamic context discussed in this section and would be better placed in Sect. 10.4.

10.5.3 Chromophore Complexation by Cyclodextrins

As for many phosphorescent probes, emission by 1-bromo-naphthalene BN is quenched by oxygen dissolved in solution. This phenomenon is nevertheless reduced by enclosing the chromophore in a basket comprising a cavity molecule and a lid molecule, namely a cyclodextrin CD and an alcohol C, respectively. When there is an excess of the latter, the shift in the $[\mathrm{CD}\text{•}\mathrm{C}] + \mathrm{BN} \rightleftarrows [\mathrm{CD}\text{•}\mathrm{BN}\text{•}\mathrm{C}]$ equilibrium with temperature can be used to build a molecular thermometer emitting light green when cold [156–158]. Apart from the emission quantum yield, the phosphorescence lifetime can also be used for thermometry, as decay is faster in contact with oxygen [158]. Interestingly, similar measurements carried out on a system in which C and CD are both cyclodextrins have demonstrated the complexity that a relatively long τ_ϕ ($\sim$ ms) can introduce into photophysical phenomena [159]. Indeed, depending on the experimental conditions, it is not unlikely that, when an initially encapsulated probe has gone into the triplet state, the complex dissociates and quenching by oxygen occurs before phosphorescence emission. Consequently, it then becomes difficult to speak of a thermodynamic thermometer in the strict sense of the term. Naturally, this type of problem does not arise when fluorescence measurements are being made, since then τ_χ is larger than τ_ϕ, and one can be sure that the collected signal will faithfully reflect the proportions of the various reagents at equilibrium. This has been demonstrated, for example, with a thermometer in which a naphthalene moiety was covalently bound to a cyclodextrin; inclusion of the chromophore

in the cavity at low temperature was shown to increase the fluorescence quantum yield [160].

Still using complexation by a cyclodextrin but exploiting other photophysical principles, a molecular thermometer was also designed using 4-(N,N-diethylamino) benzonitrile (DMABN) [161, 162]. This fluorophore has a TICT state, as described in Sect. 10.4.1, and can thus emit either at 350 nm from the Franck–Condon state FC^*, or at 530 nm from the twisted state TW^*. Moreover, in both cases, the fluorescence quantum yield is higher for the complexed form than for the free form and the formation of [FC^*•CD] is exothermic, while the formation of [TW^*•CF] is isoenthalpic. This means that a ratiometric determination of T is possible by measuring the ratio of emission intensities at 530 nm and 350 nm. Note, however, that no accurate measurement of τ_χ or τ_ϕ was made, so once again it is difficult to say whether the emission properties do actually reflect a state of thermodynamic equilibrium.

10.5.4 Acid–Base Reactions

The emission of many fluorophores with acid–base properties varies under the influence of protonation [100, 101, 163]. Consequently, as well as serving as pH indicators, these molecules can also be used as molecular thermometers. Indeed, one only has to consider the temperature dependence of the protonation equilibrium of the probe P, coupled to that of the buffer B, viz.,

$$\mathrm{PH^+} \rightleftarrows \mathrm{P} + \mathrm{H^+}\,, \qquad \mathrm{BH^+} \rightleftarrows \mathrm{B} + \mathrm{H^+}\,, \tag{10.19}$$

which is described by

$$\begin{cases} \dfrac{\mathrm{d}\ln\left([\mathrm{P}]_{\mathrm{eq}}/[\mathrm{PH^+}]_{\mathrm{eq}}\right)}{\mathrm{d}T} = \dfrac{\Delta_{\mathrm{r}}H_{\mathrm{P}}}{RT^2} + \dfrac{\mathrm{dpH}}{\mathrm{d}T}\ln 10\,, \\ \dfrac{\mathrm{d}\ln\left([\mathrm{B}]_{\mathrm{eq}}/[\mathrm{BH^+}]_{\mathrm{eq}}\right)}{\mathrm{d}T} = \dfrac{\Delta_{\mathrm{r}}H_{\mathrm{B}}}{RT^2} + \dfrac{\mathrm{dpH}}{\mathrm{d}T}\ln 10\,. \end{cases} \tag{10.20}$$

Close to the buffer $\mathrm{pk_a}$ and when its concentration is sufficient, one has $[\mathrm{B}]_{\mathrm{eq}} \approx [\mathrm{BH^+}]_{\mathrm{eq}}$, whence [148]

$$\frac{\mathrm{d}\ln\left([\mathrm{P}]_{\mathrm{eq}}/[\mathrm{PH^+}]_{\mathrm{eq}}\right)}{\mathrm{d}T} \approx \frac{\Delta_{\mathrm{r}}H_{\mathrm{P}} - \Delta_{\mathrm{r}}H_{\mathrm{B}}}{RT^2}\,. \tag{10.21}$$

Design of the thermometer depends not only on the choice of probe, but also on the choice of an ad hoc buffer, i.e., one whose acid–base properties change appropriately with temperature [164, 165]. Among the most widely used molecules are fluorescein [166–168] and its derivatives, e.g., BCECF [169] and SNARF [166], and carboxytetramethylrhodamine (TAMRA) [169]. These are diluted in the Tris buffer and the observable used is the fluorescence intensity. Note, however,

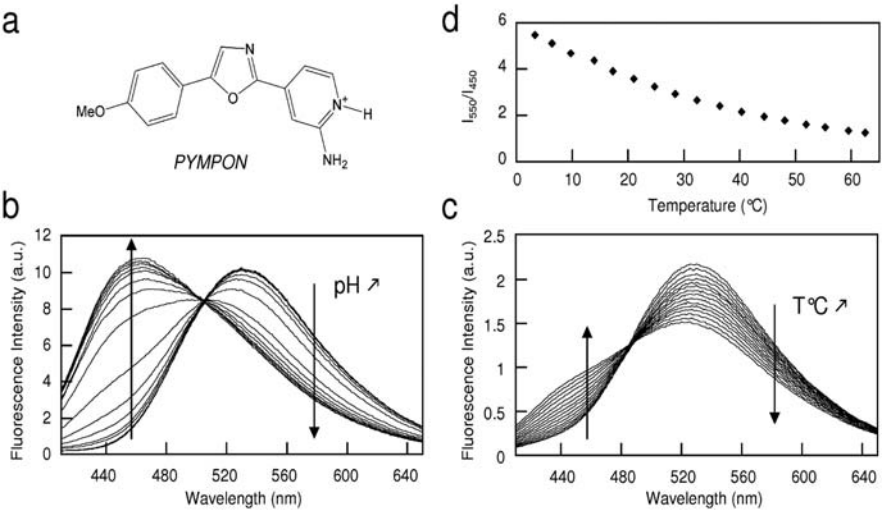

Fig. 10.12 (**a**) Molecular structure of the protonated form of PYMPON and (**b**) pH dependence of its fluorescence spectrum over the pH range 3.4–9.1 (solution at 100 nM in Britton–Robinson buffers, $T = 25°C$, $\lambda_{exc} = 339$ nm). (**c**) and (**d**) For the same compound and for a temperature ranging over 3.5–62.5°C, fluorescence spectrum and ratio of fluorescence emission intensities at 550 nm and 450 nm (associated respectively with the acidic and basic states of the probe). Solution at 100 nM in 5 mM sodium phthalate/sodium diphthalate buffer, pH = 5.2 at 25°C, $\lambda_{exc} = 341$ nm

that measurements have also been made using absorption with phenol red [170], fluorescence correlation spectroscopy (FCS) with a naturally luminescent protein [171], and two-colour fluorescence emission with a donor–acceptor pH probe (see Fig. 10.12) [148, 172, 173].

10.5.5 Modification of the Coordination Sphere of Metallic Ions

Absorption by transition metal complexes depends on the nature of the ligands and their number. Thermochromic probes have thus been designed to exploit reactions arising in the coordination sphere, whose equilibrium is shifted by a temperature change [62, 63]. In order to apply a similar strategy while continuing to use the much more sensitive fluorescence spectroscopy, a naphthalene reporter group was bound to a cryptand containing an Ni^{2+} ion. A drop in temperature causes two solvent molecules to bind onto the nickel, and concomitant fluorescence emission quenching is observed [174].

10.5.6 Thermalisation Between Excited States. General Features and Examples

Some organic probes in solution, and some rare earth ions in solid matrices, have photoluminescence spectra with relatively well resolved bands. This is the signature of thermal population, from the initially excited electronic state, of different quantum states that are clearly separated in energy. In the case of organic molecules, these are vibrational sublevels [100, 101], whereas in the case of rare earth ions, they are f–f quantum states (see Chap. 17) [67, 175]. As the thermalisation process occurs in 1–100 ps, it is much faster than emission ($\tau_\phi \sim 0.1$–100 ns). As before, the emission spectrum thus reflects a thermodynamic equilibrium, although the chemical reactions have now simply been replaced by quantum transitions:

$$\mathrm{P}_0^* \rightleftarrows \mathrm{P}_1^* \rightleftarrows \mathrm{P}_2^* \rightleftarrows \dots \,. \tag{10.22}$$

The ratio $\mathscr{R}$ of the populations of two excited sublevels n and $n+1$ separated by an energy gap $\Delta E = E_{n+1} - E_n$ is now given by the Boltzmann law:

$$\mathscr{R} = \frac{[\mathrm{P}_{n+1}^*]}{[\mathrm{P}_n^*]} = \exp\left(-\frac{\Delta E}{RT}\right) . \tag{10.23}$$

Since the intensity of each band in the spectrum is roughly proportional to the occupation number of the associated quantum state, it is easy to measure $\mathscr{R}$ and thereby deduce T. The relative sensitivity of the resulting thermodynamic thermometer is then given by

$$\mathscr{S} = \frac{\mathrm{d}\ln\mathscr{R}}{\mathrm{d}T} = \frac{\Delta E}{RT^2} . \tag{10.24}$$

It is important to choose bands a few $k_\mathrm{B}T$ apart so that they are well resolved, and a sufficient sensitivity can be obtained. However, an upper value for ΔE would be $\sim$ 24 kJ/mol. Indeed, it ensures that the upper level is not too depopulated, and thereby facilitate band detection [175].

From a practical point of view, organic dyes which spectra display a band structure are often rigid, planar molecules, soluble in an organic solvent. Emission from the higher vibrational levels of the excited state is called blue-edge fluorescence. For temperature measurements, one uses either the wavelength shift of the maximum of a broad peak [176], or the ratio of the fluorescence emissions corresponding to the 'cold' (red) and 'hot' (blue) parts of the spectrum [176, 177]. Concerning matrices doped with rare earth ions, also called thermographic phosphors, many different compositions have been tested since the first temperature measurement carried out on a Y_2O_2S ceramic doped with Eu^{3+} [178]. The most commonly used ions are Nd^{3+}, Pr^{3+}, Yb^{3+}, Er^{3+} [175] dissolved in a matrix made of silica or some ceramic, and the material may be in the form of an optical fibre [179] or a nanoparticle powder ($\sim$ 4–800 nm) [180, 181]. The ratiometric measurement at two wavelengths is the most widely used, and temperatures up to a thousand degrees can be determined.

Incidentally, measurements were also based on the wavelength shift in the position of a large band [67].

To end this section, note that the thermal population of sublevels of the ground state can also be used in thermometry. This is the case when one is interested in variations of the red-edge absorbance of dyes displaying a band spectrum [44], or when one is measuring the change in the ratio of Stokes and anti-Stokes lines in Raman spectroscopy (see Sect. 10.7.1) [191].

10.6 Procedures for Fluorescence Microscopy

In the last two sections, we have discussed the influence of temperature on the emission properties of LMTs. However, this is not yet sufficient for setting up a reliable protocol for thermometry. Indeed, the collected signal will also depend on many other experimental parameters, e.g., the optical arrangement or the nature of the sample, that need to be analysed in detail.

10.6.1 Single-Wavelength Intensity Measurement

Many thermometric measurements within labs-on-chips are made using molecules whose emission quantum yield decreases with temperature, e.g., complexes of Ru^{2+} [123–125], or fluoresceins [167–169]. Among these, rhodamine B is without doubt the most widely used [103–106, 182]. This commercial and non-toxic probe is highly sensitive to temperature, while remaining only slightly dependent on other physicochemical parameters, e.g., the pH (see Sect. 10.4.1). It is generally diluted to 0.1–1 mM in the aqueous buffer to be characterised, but it can also be integrated into one of the polymer materials making up the device [102, 183]. The thermography is then no longer limited to fluidic channels, and conversely, some measurement artifacts can be avoided.

As an example of the latter, it is rather difficult to keep the probe concentration uniform when in solution in a microchannel. Apart from trivial problems such as adsorption on the walls [183], much more subtle phenomena can arise, leading to local depletion or enrichment in fluorophore. For one thing, any temperature gradient necessarily generates a concentration gradient via the Soret effect [169]. For another, thermal inhomogeneities are also liable to modify the electrophoretic mobility field of charged probes (either directly, or via some other parameter such as the pH) and focusing phenomena may result (see Fig. 10.13) [184, 185].

With the gradual photodestruction of the fluorophore and temporal fluctuations in the intensity of the illumination source, the phenomena mentioned above continually modify the collected signal. It is therefore difficult to attribute all the observed variations in the fluorescence intensity to changes in temperature.

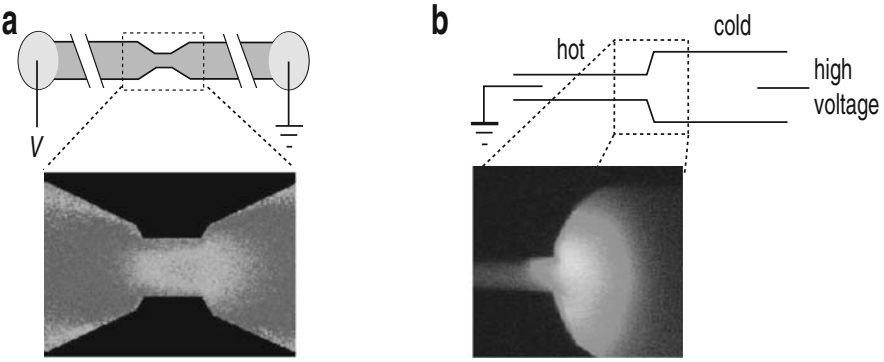

Fig. 10.13 (**a**) Visualisation of Joule heating induced by a microfluidic constriction. The concentration of the molecular thermometer, here rhodamine B, is assumed constant. From [104]. (**b**) Local accumulation of a fluorescent molecule, Oregon Green 488 carboxylic acid, obtained by thermal gradient focusing in front of a microfluidic constriction. From [185]

10.6.2 Ratiometric Intensity Measurement

Adding a second fluorescent probe whose emission intensity does not vary with temperature, e.g., rhodamine 110, carboxyfluorescein, has been envisaged as a way to correct the temperature measurement for a certain number of experimental errors [185, 186]. Indeed, the intensity of the signal associated with this new fluorophore, excited with the same light source, at the same wavelength, but not emitting at the same wavelength, then serves as an internal reference and is used to correct the thermometer response for spatiotemporal variations in the illumination, and concentration inhomogeneities due to uneven device filling. However, since the two dyes have different physicochemical properties, it is not clear that they will respond similarly to the Soret effect, or to adsorption on the channel walls. The ideal situation would thus be to rely on a probe delivering a signal whose value does not depend on its concentration.

As discussed earlier, excimer systems [23, 64, 137], dyes with band spectra [176,177], and rare-earth-doped ceramics [67,175] can be used to measure the temperature by taking the ratio of the emission intensities at two different wavelengths. We then obtain an observable with no dependence on the concentration. Unfortunately, these probes are difficult to dissolve in aqueous solutions and it is therefore important to develop new product lines devoted to on-chip thermometry. One idea is inspired by fluorescent pH probes [101,163]. Indeed, in vivo measurement of the H^+ concentration has had to face similar problems to those encountered in thermometry today [187]. Ratiometric pH meters have been developed and, by choosing a suitable buffer, they can be transformed into thermometers (see Sect. 10.5.4) [148,172,187]. Independently, a second alternative is to use molecular beacons in FRET mode to devise thermometers emitting at two different wavelengths (see Sect. 10.5.2) [148].

To end here, note that the cost in terms of instrumentation in going from monochromatic acquisition to two-colour acquisition is very modest. A simple arrangement of mirrors and dichroic filters can separate the two wavelengths and reconstruct the two corresponding images on the same CCD sensor [188].

10.6.3 Lifetime Measurement

As discussed earlier, reductions in luminescence lifetime and emission quantum yield often go hand in hand. As the former observable does not depend on the probe concentration, it can usefully replace intensity measurements [99, 132, 155, 189]. However, τ_ϕ can be very small and the cost in instrumentation soon becomes prohibitive, especially if one intends to image a whole temperature field. Indeed, if $\tau_\phi \sim$ ns, one must work in the time domain and hence use a pulsed femtosecond laser coupled to a scanner for excitation and fast detectors for signal collection [190]. If on the other hand $\tau_\phi \sim$ ms, lifetime imaging gets easier, because it is then possible to carry out measurements in the frequency domain with a modulated light source and lock-in acquisition [122].

The independence of the characteristic decay time from the concentration is not unique to photophysical phenomena. For example, in the context of isomerisation or pseudo-first order reactions, the relaxation time of the chemistry has the same property. As a result, this phenomenon has also been exploited to design molecular temperature probes. More precisely, just as luminescence could be used to probe a chemical equilibrium in Sect. 10.5, it can also be used to measure τ_χ, provided that the latter is much greater than τ_ϕ. As the kinetic constants depend on the temperature through the Arrhenius law, it was thus possible to base a molecular thermometer on a protonation reaction leading to quenching of the fluorescence of a protein [171]. In this case, the lifetime of the chemical species was measured by FCS, but the various techniques described above can also be implemented. For instance, molecular tagging thermometry is based on lifetime imaging of phosphorescent [cyclodextrin•1-bromo-naphthalene•alcohol] complexes (see Sect. 10.5.3) after localised light excitation localised, both space and time [158]. The spatial distribution of the luminescent tags then provides information about their transport by advection, while the decay of the emitted signal informs on the temperature of the carrier fluid.

10.7 Other Forms of Spectroscopy for Probing Thermodynamic Equilibria

Finally, we shall discuss two spectroscopic methods which, although not very sensitive, can be used in thermometry because they directly exploit the bulk properties of the solvent and thus do not require addition of a diluted probe.

10.7.1 Raman Spectroscopy

When a photon interacts with a molecule, it can be scattered either at the excitation frequency ν_{exc} or at a different frequency ν_{em}. One speaks of Rayleigh scattering and Raman scattering, respectively. In the second case, which is much less common than the first, each vibrational transition associated with an energy gap $h\nu_n$ results in two peaks: the Stokes peak at $\nu_{\text{em}} = \nu_{\text{exc}} + \nu_n$, corresponding to a virtual absorption departing from the ground state and a relaxation leading to the excited state, and the anti-Stokes peak at $\nu_{\text{em}} = \nu_{\text{exc}} - \nu_n$, corresponding to the opposite configuration. The ratio of the two signals is proportional to the ratio of the populations of the two vibrational levels, so it is governed by Boltzmann statistics

$$\frac{I_{\text{Stokes}}}{I_{\text{anti-Stokes}}} \propto \exp \frac{h\nu_n}{k_{\text{B}}T} , \tag{10.25}$$

and can thus be used in thermometry (see Sect. 10.5.6) [45]. Given the underlying principle, this technique is better suited to combustion temperatures than those to be observed in biological media, the anti-Stokes signal being very weak close to room temperature. However, an accuracy of ± 2 K has been obtained on cystine over the range 0–100°C [191].

In order to get round this problem of sensitivity, another Raman thermometry technique has been developed, using the Stokes line of OH bond stretching in water (see Fig. 10.14a) [192–195]. This peak in fact corresponds to 5 vibrational transitions involving two different species, one where the proton is participating in a hydrogen bond with another solvent molecule and the other where it is free [196, 197]. This equilibrium obeys the Van't Hoff law [192, 195, 197], and the temperature inside a microchannel can thus be obtained by ratiometric measurements (see Sect. 10.5). Furthermore, it should be observed that the relation $\tau_\phi \ll \tau_\chi$ is still satisfied, Raman scattering having a characteristic time of femtosecond order, whereas the lifetime of hydrogen bonds lies in the picosecond range.

10.7.2 Nuclear Magnetic Resonance

NMR thermometry arose in the context of medical imaging, and in particular, as a way of monitoring the effects of thermotherapy (see Chap. 11) [199, 200]. Several observables have been investigated, e.g., chemical shift, T_1 relaxation time, and diffusion coefficient, the signal being collected either from the solvent, i.e., water, or from a label. With the advent of miniaturised probes, i.e., microcoils, it became possible to characterise the fluids circulating in a capillary by NMR, and for example to measure the Joule heating occurring during electrophoresis (see Fig. 10.14b) [201, 202]. This determination exploits the variation in the chemical shift of the water proton with changes in temperature. Indeed, as discussed in Sect. 10.7.1, the latter may or may not be involved in a hydrogen bond, and this allows for different levels of

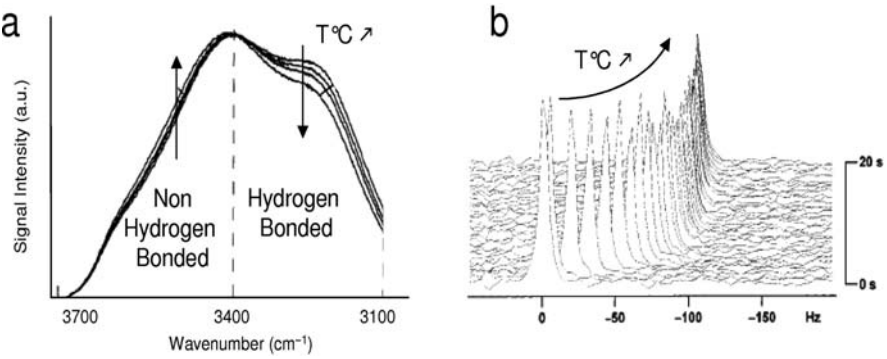

Fig. 10.14 (**a**) Temperature dependence of the Raman spectrum of water when the temperature ranges over 25–45°C ($\lambda_{exc} = 532$ nm). Taken from [192]. The band corresponding to OH bond stretching can be divided into two parts, depending on whether or not the proton is involved in a hydrogen bond. (**b**) NMR spectrum of water, as acquired when monitoring Joule heating in an electrophoresis capillary. At $t = 0$, a field of 31.6 Vm^{-1} is applied to a 100 μm diameter tube filled with 50 mM borate buffer, and the temperature increases to its equilibrium value. Taken from [201]

descreening. The transition between the free and bound forms is very fast compared with the characteristic time for signal acquisition, so the observed chemical shift is the average of the chemical shifts of the two species, weighted by their proportions at equilibrium [203].

As a conclusion, we may note that, despite the power of this method, NMR remains outside the world of the microlaboratory, doubtless due to the high cost and bulkiness of the equipment. However, this could change with the advent of cheaper and more convenient apparatus. Indeed, there will soon be no need to place the detection coil and microdevice in the air gap of a conventional NMR setup, and it will be possible to integrate decoding on-chip and carry it out in zero field [204, 205].

10.8 Conclusion and Prospects

Temperature is a fundamental parameter in Chemistry. As a consequence, various principles have been implemented to devise molecular probes, e.g., kinetic control of photophysical relaxation processes, and displacement of thermodynamic equilibrium. In this chapter, we have attempted to use reactivity as organising principle, even though the intricacy of the phenomena has sometimes made this approach look rather artificial. Indeed, luminescence emission and chemical relaxation are not always totally decorrelated [64, 138, 155, 159, 162, 171]. We nevertheless hope to have provided a useful discussion of the way molecular thermometers can be used to study microlaboratories, given that a molecular probe must always be considered alongside the microscopic technique required to observe it.

10.8.1 Fluorescent Probes

Only a few of the LMTs discussed here are currently used to characterise labs-on-chips. Examples are dyes derived from rhodamine and fluorescein, and also the complex $Ru(bpy)_3^{2+}$. The reason for their popularity probably comes from the fact that they are soluble in water and commercially available. Unfortunately, few of them can measure the temperature independently of the probe concentration and future research effort will have to focus on the design of ratiometric thermometers. Thanks to their flexibility, molecular beacons used in FRET mode may provide a good starting point. Indeed, sequences and fluorophores can be varied at will in order to obtain the desired photophysical and thermometric properties (some DNA structures are so versatile that the probe performance can simply be tuned by modifying the buffer [206]). However, because of their relatively slow chemical relaxation time, ≥ 1 μs [207], nucleic acids will not be suitable for all applications. Extremely fast systems with diffusion-limited kinetics will thus need to be developed, perhaps from existing ratiometric pH probes.

10.8.2 Optical Microscopy Techniques

As two-photon and confocal fluorescence microscopies become more widespread, 3D imaging with submicron resolution will soon be commonplace [125, 190]. Likewise, generalisation of fluorescence lifetime imaging [122, 190, 208] should lead to lifetime decay becoming a more common observable. In this context, the technique known as molecular tagging thermometry offers several interesting innovations, one being the simultaneous imaging of the temperature and velocity fields in liquids (see Sect. 10.6.3), and another being the use of phosphorescence which, with its longer lifetime, simplifies signal acquisition.

Acknowledgements This review was written with the support of the grants ANR blanche 2006 'T-wave' and ANR PNANO 2007 'Nanothermofluo' of the French Ministry of Research and Technology. We would like to thank the following colleagues for helpful suggestions regarding the manuscript and for the communication of experimental results: L. Lacroix (Muséum National d'Histoire Naturelle, Paris, France), L. Aigouy (Ecole Supérieure de Physique et Chimie Industrielles, Paris, France), M. Mortier (Ecole Nationale Supérieure de Chimie, Paris, France), T. Barilero, T. Le Saux, and L. Jullien (Ecole Normale Supérieure, Paris, France). We are also grateful to O.S. Wolfbeis (University of Regensburg, Germany), P. Guenoun (CEA, Saclay, France), B.J. Kim (University de Tokyo, Japan), M.M. Martin and D. Laage (Ecole Normale Supérieure, Paris, France) for constructive discussions.

References

1. P. Campbell (Ed.): Lab on a Chip, Nature Insight, Nature **442**, 367–418 (2006)
2. P. Tabeling: *Introduction to Microfluidics* (Oxford University Press, Oxford, 2005)
3. A. Manz, N. Graber, H.M. Widmer: Sens. Actuators B **1**, 244 (1990)
4. S.C. Terry, J.H. Jerman, J.B. Angell: IEEE Trans. Electron. Devices ED **26**, 1880 (1979)
5. M.A. Burns, B.N. Johnson, S.N. Brahmasandra, K. Handique, J.R. Webster, M. Krishnan, T.S. Sammarco, P.M. Man, D. Jones, D. Heldsinger, C.H. Mastrangelo, D.T. Burke: Science **282**, 484 (1998)
6. C.J. Easley, J.M. Karlinsey, J.M. Bienvenue, L.A. Legendre, M.G. Roper, S.H. Feldman, M.A. Hughes, E.L. Hewlett, T.J. Merkel, J.P. Ferrance, J.P. Landers: Proc. Natl. Acad. Sci. USA **103**, 19272 (2006)
7. N.J. Petersen, R.P.H. Nikolajsen, K.B. Mogensen, J.P. Kutter: Electrophoresis **25**, 253 (2004)
8. A. Manz, J.C. Eijkel: Pure Appl. Chem. **73**, 1555 (2001)
9. W.A. Gobie, C.F. Cornelius: J. Chromatogr. **516**, 191 (1990)
10. N.J. Dovichi, J. Zhang: Angew. Chem. Int. Ed. **39**, 4463 (2000)
11. S.C. Jacobson, C.T. Culbertson, J.E. Daler, J.M. Ramsey: Anal. Chem. **70**, 3476 (1998)
12. A.J. de Mello: Lab Chip **1**, 24N (2001)
13. P.-A. Auroux, Y. Koc, A. deMello, A. Manz, P.J.R. Day: Lab Chip **4**, 534 (2004)
14. C. Zhang, J. Xu, W. Ma, W. Zheng: Biotechnol. Adv. **24**, 243 (2006)
15. M.A. Northup, M.T. Ching, R.M. White, R.T. Watson: In: Transducers 1993. Digest of the Technical Papers of the Seventh International Conference on Solid State Sensors and Actuators (1993) pp. 924–926
16. P. Wilding, M.A. Shoffner, L.J. Kricka: Clin. Chem. **40**, 1815 (1994)
17. J. Khandurina, T.E. McKnight, S.C. Jacobson, L.C. Waters, R.S. Foote, J.M. Ramsey: Anal. Chem. **72**, 2995 (2000)
18. M.G. Roper, C.J. Easley, L.A. Legendre, J.A.C. Humphrey, J.P. Landers: Anal. Chem. **79**, 1294 (2007)
19. E.T. Lagally, C.A. Emrich, R.A. Mathies: Lab Chip **1**, 102 (2001)
20. R.H. Liu, J. Yang, R. Lenigk, J. Bonanno, P. Grodzinski: Anal. Chem. **76**, 1824 (2004)
21. H. Nakano, K. Matsuda, M. Yohda, T. Nagamune, I. Endo, T. Yamane: Biosci. Biotech. Biochem. **58**, 349 (1994)
22. M.U. Kopp, A.J. de Mello, A. Manz: Science **280**, 1046 (1998)
23. K. Sun, A. Yamaguchi, Y. Ishida, S. Matsuo, H. Misawa: Sens. Actuators B **84**, 283 (2002)
24. J. Liu, M. Enzelberger, S. Quake: Electrophoresis **23**, 1531 (2002)
25. M. Krishnan, V.M. Ugaz, M.A. Burns: Science **298**, 793 (2002)
26. D. Braun, N.L. Goddard, A. Libchaber: Phys. Rev. Lett. **91**, 158103 (2003)
27. D. Braun: Mod. Phys. Lett. B **18**, 775 (2004)
28. M. Henning, D. Braun: Appl. Phys. Lett. **87**, 183901 (2005)
29. S. Leavitt, E. Freire: Curr. Opin. Struct. Biol. **11**, 560 (2001)
30. P.C. Weber, F.R. Salemme: Curr. Opin. Struct. Biol. **13**, 115 (2003)
31. F.E. Torres, P. Kuhn, D. De Bruyker, A.G. Bell, M.V. Wolkin, E. Peeters, J.R. Williamson, G.B. Anderson, G.P. Schmitz, M.I. Recht, S. Schweizer, L.G. Scott, J.H. Ho, S.A. Elrod, P.G. Schultz, R.A. Lerner, R.H. Bruce: Proc. Natl. Acad. Sci. USA **101**, 9517 (2004)
32. E.A. Johannessen, J.M.R. Weaver, L. Bourova, P. Svoboda, P.H. Cobbold, J.M. Cooper: Anal. Chem. **74**, 2190 (2002)
33. E.B. Chancellor, J.P. Wikso, F. Baudenbacher, M. Radparvar, D. Osterman: Appl. Phys. Lett. **85**, 2408 (2004)
34. Y. Zhang, S. Tadigadapa: Biosens. Bioelectron. **19**, 1733 (2004)
35. J. Lerchner, A. Wolf, G. Wolf, I. Fernandez: Thermochim. Acta **446**, 168 (2006)
36. T.K. Hakala, J.J. Toppari, P. Törmä: J. Appl. Phys. **101**, 034512 (2007)
37. J. Xu, R. Reiserer, J. Tellinghuisen, J.P. Wikswo, F.J. Baudenbacher: Anal. Chem. **80**, 2728 (2008)
38. M. Volk: Eur. J. Org. Chem. **2605** (2001)

39. R. Callender, R.B. Dyer: Chem. Rev. **106**, 3031 (2006)
40. M. Eigen, L. de Mayer: In: *Techniques of Organic Chemistry*, Vol. 8, 2nd edn., ed. by S.L. Friess, E.S. Lewis, A. Wiessberger (Wiley Interscience, New York, 1963) pp. 895–1054
41. C.R. Cantor, P.R. Schimmel: *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, Freeman, New York (1980)
42. T.C. French, G.G. Hammes: Methods Enzymol. **16**, 3 (1969)
43. H. Ma, C. Wan, A.H. Zewail: J. Am. Chem. Soc. **128**, 6338 (2006)
44. A. Seilmeier, P.O.J. Scherer, W. Kaiser: Chem. Phys. Lett. **105**, 140 (1984)
45. P.R.N. Childs, J.R. Greenwood, C.A. Long: Rev. Sci. Instrum. **71**, 2959 (2000)
46. P. Löw: High spatial resolution fluorescent thermometry: Thermal characterization of sub-micrometer structures in dry and liquid conditions (PhD Thesis, Paul Sabatier University, Toulouse, 2008)
47. H.F. Arata, Y. Rondelez, H. Noji, H. Fujita: Anal. Chem. **77**, 4810 (2005)
48. X. Zhang, H. Choi, A. Datta, X. Li: J. Micromech. Microeng. **16**, 900 (2006)
49. D. Chu, W.-K. Wong, K.E. Goodson, R.F.W. Pease: J. Vac. Sci. Technol. B **21**, 2985 (2003)
50. M.E. Bourg, W.E. van der Veer, A.G. Grüell, R.M. Penner: Nanolett. **7**, 3208 (2007)
51. T. Yamamoto, T. Nojima, T. Fujii: Lab Chip **2**, 197 (2002)
52. C.F. Chou, R. Changrani, P. Roberts, D. Sadler, J. Burdon, F. Zenhausern, S. Lin, A. Mulholland, N. Swami, R. Terbrueggen: Microelectron. Eng. **61–62**, 921 (2002)
53. O. Fudym, C. Pradère, J.-C. Batsale: Chem. Eng. Sci. **62**, 4054 (2007)
54. K. Swinney, D. Markov, D. J. Bornhop: Anal. Chem. **72**, 2690 (2000)
55. K. Swinney, D. J. Bornhop: Electrophoresis **23**, 613 (2002)
56. C.J. Easley, L.A. Legendre, M.G. Roper, T.A. Wavering, J.P. Ferrance, J.P. Landers: Anal. Chem. **77**, 1038 (2005)
57. P.M. Celliers, J. Conia: Appl. Opt. **39**, 3396 (2000)
58. D. Boyer, P. Tamarat, A. Maali, B. Lounis, M. Orrit: Science **297**, 1160 (2002)
59. N.J. Dovichi, J.M. Harris: Anal. Chem. **51**, 728 (1979)
60. K. Uchiyama, A. Hibara, H. Kimura, T. Sawada, T. Kitamori: Jpn. J. Appl. Phys. **39**, 5316 (2000)
61. M. Tokeshi, M. Uchida, A. Hibara, T. Sawada, T. Kitamori: Anal. Chem. **77**, 1038 (2005)
62. O. Sano, R. Takaki: J. Phys. Soc. Jpn **52**, 1549 (1983)
63. H. Wätzig: Chromatographia **33**, 445 (1992)
64. J. Lou, T.M. Finegan, P. Mohsen, T.A. Haton, P.E. Laibinis: Rev. Anal. Chem. **18**, 235 (1999)
65. S. Uchiyama, A.P. de Silva, K. Iwai: J. Chem. Ed. **83**, 720 (2006)
66. N. Chandrasekharan, L.A. Kelly: In: *Reviews in Fluorescence*, Vol. 1, ed. by C.D. Geddes, J.R. Lakowicz (Kluwer Academic/Plenum, New York, 2004) pp. 21–40
67. S.W. Allison, G.T. Gillies: Rev. Sci. Instrum. **68**, 2615 (1997)
68. P.G. de Gennes, J. Prost: *The Physics of Liquid Crystals*, 2nd edn. (Oxford University Press, Oxford, 1993)
69. V. Hohreiter, S.T. Wereley, M.G. Olsen, J.N. Chung: Meas. Sci. Technol. **13**, 1072 (2002)
70. K. Chung, J.K. Cho, E.S. Park, V. Breedveld, H.Lu: Anal. Chem. **81**, 991 (2009)
71. C.B. Müller, K. Weiss, A. Loman, W. Richtering: Lab Chip **9**, 1248 (2009)
72. A.M. Chaudhari, T.M. Woudenberg, M. Albin, K.E. Goodson: J. Microelectromechanical Systems **7**, 345 (1998)
73. A. Iles, R. Fortt, A.J. de Mello: Lab Chip **5**, 540 (2005)
74. J. Noh, S.W. Sung, M.K. Jeon, S.H. Kim, L.P. Lee, S.I. Woo: Sens. Actuators A **122**, 196 (2005)
75. L. Liu, S. Peng, W. Wen, P. Sheng: Appl. Phys. Lett. **91**, 093513 (2007)
76. F. Gillot, F.O. Morin, H.F. Arata, R. Guégan, H. Tanaka, H. Fujita: Lab Chip **7**, 1600 (2007)
77. J.L. Fergason: Appl. Opt. **7**, 1729 (1968)
78. *Handbook of Thermochromic Liquid Crystal Technology* (Hallcrest, Glenview, 1991)
79. C.R. Smith, D.R. Sabatino, T.J. Praisner: Exp. Fluids **30**, 190 (2001)
80. C.E. Stephens, F.N. Sinnadurai: J. Phys. E: Sci. Instrum. **7**, 641 (1974)
81. G. Aszódi, J. Szabon, I. Jánossy, V. Székely: Solid-State Electron. **24**, 1127 (1981)
82. A. Csendes, V. Székely, M. Rencz: Microelectron. Eng. **31**, 281 (1996)

83. H.G. Schild: Prog. Pol. Sci. **17**, 163 (1992)
84. V.A. Baulin, A. Halperin: Macromolecules **35**, 6432 (2002)
85. C. Wu, X. Wang: Phys. Rev. Let. **80**, 4092 (1998)
86. E.I. Tiktopulo, V.E. Bychkova, J. Ricka, O.B. Ptitsyn: Macromolecules **27**, 2879 (1994)
87. A.V. Gorelov, A. Du Chesne, K.A. Dawson: Physica A **240**, 443 (1997)
88. S. Fery-Forgues, J.-P. Fayet, A. Lopez: Photochem. Photobiol. A **70**, 229 (1993)
89. S. Uchiyama, Y. Matsumura, A.P. de Silva, K. Iwai: Anal. Chem. **75**, 5926 (2003)
90. S. Uchiyama, Y. Matsumura, A.P. de Silva, K. Iwai: Anal. Chem. **76**, 1793 (2004)
91. F.M. Winnik: Macromolecules **23**, 233 (1990)
92. R. Nomura, K. Yamada, T. Masuda: Chem. Comm. **478** (2002)
93. C. Mangeney, F. Ferrage, I. Aujard, V. Marchi-Artzner, L. Jullien, O. Ouari, E.D. Rékai, A. Laschewsky, I. Vikholm, J.W. Sadowski: J. Am. Chem. Soc. **124**, 5811 (2002)
94. J. Li, X. Hong, Y. Liu, D. Li, Y. Wang, J. Li, Y. Bai, T. Li: Adv. Mater. **17**, 163 (2005)
95. M. Karg, I. Pastoriza-Santos, J. Pérez-Juste, T. Hellweg, L.M. Liz-Marzán: Small **3**, 1222 (2007)
96. J. Lee, A.O. Govorov, N.A. Kotov: Angew. Chem. Int. Ed. **44**, 7439 (2005)
97. A.O. Govorov, G.W. Bryant, W. Zhang, T. Skeini, J. Lee, N.A. Kotov, J.M. Slocik, R.R. Naik: Nano Lett. **6**, 984 (2006)
98. Y. Liu, D.K. Cheng, G.J. Sonek, M.W. Berns, C.F. Chapman, B.J. Tromberg: Biophys. J. **68**, 2137 (1995)
99. C.F. Chapman, Y. Liu, G.J. Sonek, B.J. Tromberg: Photochem. Photobiol. **62**, 416 (1995)
100. J.R. Lakowicz: *Principles of Fluorescence Spectroscopy*, 2nd edn. (Kluwer Academic Plenum Publishers, New York, 1999)
101. B. Valeur: *Molecular Fluorescence: Principle and Applications* (Wiley-VCH, Weinheim, 2002)
102. V. Romano, A.D. Zweig, M. Frenz, H.P. Weber: Appl. Phys. B **49**, 527 (1989)
103. H. Kato, T. Nishizaka, T. Iga, K. Kinosita, S. Ishiwata: Proc. Natl. Acad . Sci. USA **96**, 9602 (1999)
104. D. Ross, M. Gaitan, L.E. Locascio: Anal. Chem. **73**, 4117 (2001)
105. D. Erickson, D. Sinton, D. Li: Lab Chip **3**, 141 (2003)
106. H.F. Arata, P. Low, K. Ishizuka, C. Bergaud, B. Kim, H. Noji, H. Fujita: Sens. Actuators B **117**, 339 (2006)
107. M.N. Slyadnev, Y. Tanaka, M. Tokeshi, T. Kitamori: Anal. Chem. **73**, 4037 (2001)
108. I. Rosenthal, P. Peretz, K.A. Muszkat: J. Phys. Chem. **83**, 350 (1979)
109. D.A. Hinckley, P.G. Seybold, D.P. Borris: Spectrochim. Acta A **42**, 747 (1986)
110. D.A. Hinckley, P.G. Seybold: Spectrochim. Acta A **44**, 1053 (1988)
111. I.L. Arbeloa, P.R. Ojeda: Chem. Phys. Lett. **79**, 347 (1981)
112. Z.R. Grabowki, K. Rotkiewicz, W. Rettig: Chem. Rev. **103**, 3899 (2003)
113. M. Vogel, W. Rettig, R. Sens, K.H. Drexhage: Chem. Phys. Lett. **147**, 452 (1988)
114. M. Vogel, W. Rettig, R. Sens, K.H. Drexhage: Chem. Phys. Lett. **147**, 461 (1988)
115. J.A.B. Ferreira, S.M.B. Costa, L.F.V. Ferreira: J. Phys. Chem. A **104**, 11909 (2000)
116. J.N. Demas, B.A. DeGraff: Coord. Chem. Rev. **211**, 317 (2001)
117. J. van Houten, R.J. Watts: J. Am. Chem. Soc. **98**, 4853 (1976)
118. S.R. Allsopp, A. Cox, T.J. Kemp, W.J. Reed: J. Chem. Soc., Faraday Trans. **174**, 1275 (1978)
119. R.S. Lumpkin, E.M. Kober, L.A. Worl, Z. Murtaza, T.J. Meyer: J. Phys. Chem. **94**, 239 (1990)
120. J.M. Lupton: Appl. Phys. Lett. **81**, 2478 (2002)
121. A. Mills, C. Tommons, R.T. Bailey, M.C. Tedford, P.J. Crilly: Analyst **131**, 495 (2006)
122. G. Liebsch, I. Klimant, S.O. Wolfbeis: Adv. Mater. **11**, 1296 (1999)
123. Y. Sato, G. Irisawa, M. Ishizuka, K. Hishida, M. Maeda: Meas. Sci. Technol. **14**, 114 (2003)
124. O. Filevich, R. Etchenique: Anal. Chem. **78**, 7499 (2006)
125. E. van Keuren, D. Littlejohn, W. Schrof: J. Phys. D Appl. Phys. **37**, 2938 (2004)
126. S.I. Weissman: J. Chem. Phys. **10**, 214 (1942)
127. M.L. Bhaumik, L.J. Nugent: J. Chem. Phys. **43**, 1680 (1965)
128. G.A. Crosby: Mol. Cryst. **1**, 37 (1966)

129. M.T. Berry, P.S. May, H. Xu: J. Phys. Chem. **100**, 9216 (1996)
130. N. Sabbatini, M. Guardigli, J.-M. Lehn: Coord. Chem. Rev. **123**, 201 (1993)
131. G.E. Khalil, K. Lau, G.D. Phelan, B. Carlson, M. Goutherman, J.B. Callis, L.R. Dalton: Rev. Sci. Instrum. **75**, 192 (2004)
132. S.M. Borisov, O.S. Wolfbeis: Anal. Chem. **78**, 5094 (2006)
133. P. Kolodner, J.A. Tyson: Appl. Phys. Lett. **40**, 782 (1982)
134. E. Van Keuren, M. Cheng, O. Albertini, C. Luo, J. Currie, M. Paranjape: Sens. Mater. **17**, 1 (2005)
135. O. Zohar, M. Ikeda, H. Shinagaw, H. Inoue, H. Nakamura, D. Elbaum, D.L. Alkon, T. Yoshioka: Biophys. J. **74**, 82 (1998)
136. N. Chandrasekharan, L.A. Kelly: J. Am. Chem. Soc. **123**, 9898 (2001)
137. H.E. Gossage, L.A. Melton: Appl. Optics **26**, 2256 (1987)
138. M.T. Albelda, E. Garcìa-España, L. Gil, J.C. Lima, C. Lodeiro, J.S. de Melo, M.J. Melo, A.J. Parola, F. Pina, C. Soriano: J. Phys. Chem. B **107**, 6573 (2003)
139. F. Pragst, H.-J. Hamann, K. Teuchner, M. Naether, W. Becker, S. Daehne: Chem. Phys. Lett. **48**, 36 (1977)
140. J. Lou, T.A. Haton, P.E. Laibinis: Anal. Chem. **69**, 1262 (1997)
141. S.W. Buckner, R.A. Forlines, J.R. Gord: Appl. Spectrosc. **53**, 115 (1999)
142. C.A. Parker: Adv. Photochem. **2**, 305 (1964)
143. J.C. Fister, D. Rank, J.M. Harris: Anal. Chem. **67**, 4269 (1995)
144. C. Baleizão, S. Nagl, S.M. Borisov, M. Schäferling, O.S. Wolfbeis, M. Berberan-Santos: Chem. Eur. J. **13**, 3643 (2007)
145. S. Tyagi, F.R. Kramer: Nat. Biotechnol. **14**, 303 (1996)
146. G. Bonnet, S. Tyagi, A. Libchaber, F.R. Kramer: Proc. Natl. Acad. Sci. USA **96**, 6171 (1999)
147. S.A.E. Marras, F.R. Kramer, S. Tyagi: Nucleic Acids Res. **30**, e122 (2002)
148. T. Barilero, T. Le Saux, C. Gosse, L. Jullien: Anal. Chem. **81**, 7988 (2009)
149. A. Dodge, G. Turcatti, I. Lawrence, N.F. de Rooij, E. Verpoorte: Anal. Chem. **76**, 1778 (2004)
150. C. Nellåker, U. Wållgren, H. Karlsson: Clin. Chem. **53**, 98 (2007)
151. F. Narberhaus, T. Waldminghaus, S. Chowdhury: FEMS Microbiol. Rev. **30**, 3 (2006)
152. R. Tashiro, H. Sugiyama: Angew. Chem. Int. Ed. **42**, 6018 (2003)
153. M. Berezovski, S.N. Krylov: Anal. Chem. **76**, 7114 (2004)
154. G. Hungerford, M. Martinez-Insua, D.J.S. Birch, B.D. Moore: Angew. Chem. Int. Ed. Engl. **35**, 326 (1996)
155. S. Jeon, J. Turner, S. Granick: J. Am. Chem. Soc. **125**, 9908 (2003)
156. W.K. Hartmann, M.H.B. Gray, A. Ponce, D.G. Nocera, P.A. Wong: Inorg. Chim. Acta **243**, 239 (1996)
157. S.L. Thomson, D. Maynes: J. Fluids Eng. **123**, 293 (2001)
158. H. Hu, M. Koochesfahani, C. Lum: Exp. Fluids **40**, 753 (2006)
159. R.E. Brewster, M.J. Kidd, M.D. Schuh: Chem. Comm. **12**, 1134 (2001)
160. H. Nakashima, Y. Takenaka, M. Higashi, N. Yoshida: J. Chem. Soc., Perkin Trans. **2**, 2096 (2001)
161. M. El Baraka, R. García, E. Quiñones: J. Photochem. Photobiol. A **79**, 181 (1994)
162. I.D. Figueroa, M. El Baraka, E. Quiñones, O. Rosario, M. Deumié: Anal. Chem. **70**, 3974 (1998)
163. R.P. Haugland: *Handbook of Fluorescent Probes and Research Products*, 9th edn. (Molecular Probes, Eugene, 2002)
164. N.E. Good, G.D. Winget, W. Winter, T.N. Connolly, S. Izawa, R.M.M. Singh: Biochemistry **5**, 467 (1966)
165. H. Fukada, K. Takahashi: Proteins **33**, 159 (1998)
166. J. Coppeta, C. Rogers: Exp. Fluids **25**, 1 (1998)
167. H. Mao, T. Yang, P.S. Cremer: J. Am. Chem. Soc. **124**, 4432 (2002)
168. G.L. Liu, J. Kim, Y. Lu, L.P. Lee: Nature Mat. **5**, 27 (2006)
169. S. Duhr, S. Arduini, D. Braun: Eur. Phys. J. E **15**, 277 (2004)
170. A.E. Straub, W.R. Seitz: Anal. Chem. **65**, 1491 (1993)
171. F.H.C. Wong, D.S. Banks, A. Abu-Arish, C. Fradin: J. Am. Chem. Soc. **129**, 10302 (2007)

172. S. Charier, O. Ruel, J.-B. Baudin, D. Alcor, J.-F. Allemand, A. Meglio, L. Jullien: Ang. Chem. **116**, 4889 (2004)
173. S. Charier, O. Ruel, J.-B. Baudin, D. Alcor, J.-F. Allemand, A. Meglio, L. Jullien, B. Valeur: Chem. Eur. J. **12**, 1097 (2006)
174. M. Engeser, L. Fabbrizzi, M. Licchelli, D. Sacchi: Chem. Comm. 1191 (1999)
175. S.A. Wade, S.F. Collins, G.W. Baxter: J. Appl. Phys. **94**, 4743 (2003)
176. K.F. Schrum, A.M. Williams, S.A. Haether, D. Ben-Amotz: Anal. Chem. **66**, 2788 (1994)
177. F. Bai, L.A. Melton: Appl. Spectroscop. **51**, 1276 (1997)
178. H. Kusama, O.J. Sovers, T. Yoshioka: Jap. J. Appl. Phys. **15**, 2349 (1976)
179. S.A. Wade, J.C. Muscat, S.F. Collins, G.W. Baxter: Rev. Sci. Instrum. **70**, 4279 (1999)
180. A.S.S. de Camargo, J.F. Possatto, L.A. de O. Nunes, E.R. Botero, E.R.M. Andreeta, D. Garcia, J.A. Eiras: Solid State Commun. **137**, 1 (2006)
181. B. Samson, L. Aigouy, R. Latempa, G. Tessier, M. Aprili, M. Mortier, J. Lesueur, D. Fournier: J. Appl. Phys. **102**, 024305 (2007)
182. R.M. Guijt, A. Dodge, G.W.K. van Dedem, N.F. de Rooij, E. Verpoorte: Lab Chip **3**, 1 (2003)
183. R. Samy, T. Glawdel, C.L. Ren: Anal. Chem. **80**, 369 (2008)
184. T. Huang, J. Pawliszyn: Electrophoresis **23**, 3504 (2002)
185. D. Ross, L.E. Locascio: Anal. Chem. **74**, 2556 (2002)
186. J. Sakakibara, R.J. Adrian: Exp. Fluids **26**, 7 (1999)
187. J. Coppeta, C. Rogers: Exp. Fluids **25**, 1 (1998)
188. K. Kinosita, H. Itoh, S. Ishiwata, K. Hirano, T. Nishizaka, T. Hayakawa: J. Cell Biol. **115**, 67 (1991)
189. S.W. Allison, G.T. Gillies, A.J. Rondinone, M.R. Cates: Nanotechnol. **14**, 859 (2003)
190. R.K.P. Benninger, Y. Koç, O. Hofmann, J. Requejo-Isidro, M.A.A. Neil, P.M.W. French, A.J. deMello: Anal. Chem. **78**, 2272 (2006)
191. B.J. Kip, R.J. Meier: Appl. Spectrosc. **44**, 707 (1990)
192. K.L. Davis, K.-L.K. Liu, M. Lanan, M.D. Morris: Anal. Chem. **65**, 293 (1993)
193. K.-L.K. Liu, K.L. Davis, M.D. Morris: Anal. Chem. **66**, 3744 (1994)
194. S.H. Kim, J. Noh, M.K. Jeon, K.W. Kim, L.P. Lee, S.I. Woo: J. Micromech . Microeng. **16**, 526 (2006)
195. J.D. Smith, C.D. Cappa, W.S. Drisdell, R.C. Cohen, R.J. Saykally: J. Am. Chem. Soc. **128**, 12892 (1996)
196. G.E. Walrafen: J. Chem. Phys. **47**, 114 (1967)
197. D.M. Carey, G.M. Korenowski: J. Chem. Phys. **108**, 2669 (1998)
198. D. Laage, J.T. Hynes: Science **311**, 832 (2006)
199. J. Olsrud, R. Wirestam, S. Brockstedt, A.M.K. Nilsson, K.-G. Tramberg, F. Ståhlberg, B.R.R. Persson: Phys. Med. Biol. **43**, 2597 (1998)
200. M. Hentschel, W. Dreher, P. Wust, S. Röll, D. Leibfritz, R. Felix: Phys. Med. Biol. **44**, 2397 (1999)
201. M.E. Lacey, A.G. Webb, J.V. Sweedler: Anal. Chem. **72**, 4991 (2000)
202. M.E. Lacey, A.G. Webb, J.V. Sweedler: Anal. Chem. **74**, 4583 (2002)
203. J.C. Hindman: J. Chem. Phys. **44**, 4582 (1965)
204. F. Verpillat, M.P. Ledbetter, S. Xu, D.J. Michalak, C. Hilty, L.-S. Bouchard, S. Antonijevic, D. Budker, A. Pines: Proc. Natl. Acad. Sci. USA **105**, 2271 (2008)
205. M.P. Ledbetter, I.M. Savukov, D. Budker, V. Shah, S. Knappe, J. Kitching, D.J. Michalak, S. Xu, A. Pines: Proc. Natl. Acad. Sci. USA **105**, 2286 (2008)
206. A. Bourdoncle, A. Estévez-Torres, C. Gosse, L. Lacroix, P. Verkhoff, T. Le Saux, L. Jullien, J.-L. Mergny: J. Am. Chem. Soc. **128**, 11094 (2006)
207. G. Bonnet, O. Krichevsky, A. Libchaber: Proc. Natl. Acad. Sci. USA **95**, 8602 (1998)
208. A.D. Elder, S.M. Matthews, J. Swartling, K. Yunus, J.H. Frank, C.M. Brennan, A.C. Fisher, C.F. Kaminsky: Opt. Express **14**, 5456 (2006)

Chapter 11
Cell Targeting and Magnetically Induced Hyperthermia

Etienne Duguet, Lucile Hardel, and Sébastien Vasseur

11.1 Introduction

11.1.1 Nanomedicine. An Application of Nanoscience and Nanotechnology

With the recent development of efficient and reproducible methods for synthesis, stable aqueous dispersions of individual particles can be prepared, in which the particle sizes can be accurately adjusted from a few nanometers to a few tens of nanometers [1]. Provided that their physical and chemical surface properties can be suitably adapted, these objects are small enough to circulate within the human body without risk of causing an embolus, since the finest capillaries (those of the lungs) have a minimal internal diameter of 5 μm. They can also escape from the blood compartment by windows of diameter around 100 nm in certain epithelia with permeability defects, such as those located in tumours and centers of infection, whereby they may then accumulate in such tissues. Furthermore, the smallest particles can migrate from the cardiovascular system into the lymph system. Finally, under the right conditions, they can enter cells and their various compartments. They should quickly become indispensable in the field of biological labelling, image contrast enhancement, the delivery of active principles, and the treatment of many different pathologies, by virtue of their novel physical properties [2, 3].

Research scientists and physicians thus have at their disposal new tools for understanding biological processes, strengthening the value of medical diagnosis, and even developing new therapeutic strategies. And so a new and largely cross-disciplinary field of investigation was born: nanomedicine, which we may define as molecular scale surveillance, repair, construction, and control of human biological systems by means of nanoscale devices [4].

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_11, © Springer-Verlag Berlin Heidelberg 2009

11.1.2 An Incomplete and Complex Set of Requirements

The development of new diagnostic and therapeutic approaches is inevitably long and tortuous. Beyond an initial set of requirements based on the knowledge of the day and strategic hypotheses that seem relevant at the time, in vitro, then in vivo, and finally clinical trials often lead to failure, whence the whole approach may need to be reconsidered, or even abandoned. Indeed, even the most active label or drug in vitro will be unusable in vivo if it cannot remain long enough in the blood compartment to reach its target, due to premature metabolism, immunological reactions, excessive toxicity, fast excretion, or unexpected capture by non-targeted tissues [5]. Unfortunately then, successful innovations are few and far between.

11.2 In Vivo Applications of Nanoparticles

Intravenous administration (injection or perfusion) is in principle the most general method for attaining a given organ or tissue, since all cells are supplied directly or indirectly by the blood system. To be administered by this channel, not only must nanoparticles not aggregate in the blood or be prematurely degraded, but they must also be able to foil the complex mechanisms of the immune system, which will naturally do everything in their power to eliminate them. Our first task is thus to spell out the mechanisms whereby nanoparticles are eliminated from the plasma, the factors influencing their biodistribution, and the solutions developed to make them stealthy with respect to the immune system. We shall then discuss ways of delivering them as specifically as possible to a given target. Finally, we shall describe current and future medical applications for nanoparticles administered intravenously, both in diagnosis and in therapeutics.

11.2.1 Nanoparticles in the Blood Compartment

It is the mononuclear phagocyte system (MPS), also known as the reticuloendothelial system, which actively engages in the extravasation of any foreign body of an infectious nature or otherwise from the blood system. It thus constitutes the first line of the immune defense system. It comprises a population of macrophage cells strategically placed throughout the body. They are mainly found in the bone marrow, where they are produced, the blood, where they circulate in the form of monocytes, the alveoli in the lungs, the spleen, and notably in the liver, where they are called Kupffer cells [6].

Their role is to recognise and eliminate senescent cells, micro-organisms, and particles from the blood compartment. In particular in the latter case, the blood clearance process is often engaged by a preliminary step known as opsonisation, in which circulating proteins (various subclasses of immunoglobulins, elements

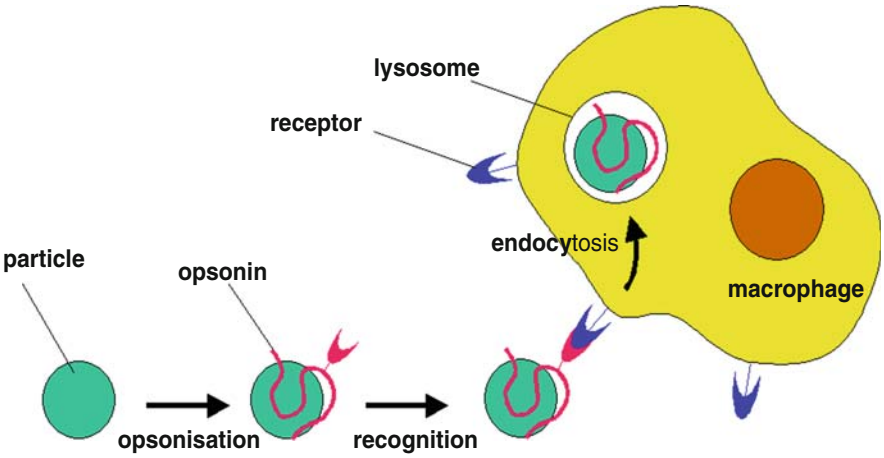

Fig. 11.1 Schematic view of the mechanisms implemented by the MPS for recognition and clearance of nanoparticles arriving accidentally in the blood compartment

of the complement, fibronectin, etc.) adsorb onto the surface of the particles (see Fig. 11.1). Labelled in this way, the particles are then easily recognised by the macrophages, which carry specific receptors. They are subsequently internalised by endocytosis and accumulate in the lysosomes, where they are eventually degraded by lysosomal enzymes. So within a few minutes, the particles are liable to be eliminated from the blood compartment and find themselves immobilised in the liver (up to 90%), the spleen, and to a lesser extent in the bone marrow [5].

This fatal destiny for the nanoparticles can nevertheless be exploited to deliver particles specifically to these organs and diagnose or treat pathologies affecting them. One then speaks of passive targeting, since it is spontaneously managed by the MPS. However, if the organ to be targeted is anything other than the liver and/or the spleen, it is essential to avoid, or at worst significantly slow down the opsonisation process. The idea then is to make the particles stealthy with respect to the MPS and thereby increase their chances of reaching the target.

11.2.2 Designing Particles with Extended Vascular Lifetime

In order to extend the plasma half-life, it has been clearly established today that the adsorption of opsonins must be hindered, or better, prevented, by playing on both the size and the surface properties of the particles [7]. For example, it has been shown that the smaller the particle radius, the longer the particles will circulate in the blood compartment, because the radius of curvature influences the kind and/or amount of opsonins adsorbed. Further, the lower the surface charge density, the longer the circulation time will be. Finally, hydrophobic particles are so quickly

eliminated from the blood compartment that it would appear that the preliminary opsonisation stage is not even necessary for their clearance. In contrast, particles with a particularly hydrophilic surface have every chance of circulating for a long time, rather like the red blood cells themselves, whose lifetime of 120 days can be explained by the presence at their surface of a hydrophilic barrier composed of oligosaccharide groups. For synthetic particles, this physicochemical concept of steric repulsion can be reproduced by coating the particle surface with flexible and hydrophilic macromolecules. Arranged rather perpendicularly to the surface, they form a brush structure which constitutes an effective steric barrier.

The macromolecules used are generally polysaccharides like dextran (resulting from the bacterial fermentation of sucrose). Other macromolecules of biological origin have been used, such as polysialic acid or heparin, but their sometimes high cost and/or risk of immunological consequences has been a hindrance in their development. Synthetic macromolecules have thus received more attention. An example is poly(ethylene glycol), known as PEG, with chemical formula $HO–(CH_2–CH_2–O)_n–H$, which is widely used in galenic pharmacy, where it is commonly conjugated with active principles (small molecules or peptides, proteins, antibodies, oligonucleotides) to reduce their immunogenicity, increase their plasma half-life (by reducing the rate of renal clearance), and hence increase also their bioavailability [5]. This method is so widely used that one speaks of pegylation. PEG can be anchored to the surface when the particles are synthesised, grafted on by one end later, or simply physisorbed. For maximal efficiency with regard to opsonisation, the molar mass of the PEG chains must lie in the range 2 000–5 000 g/mol.

In this way, the plasma half-life of so-called stealth particles can be prolonged from a few minutes to several hours. If they are small enough, they will also be able to leave the blood compartment in regions where the vascular epithelium has discontinuities. Indeed, the integrity of the endothelial barrier is often perturbed near centers of infection and certain kinds of tumour. Stealth particles then have the opportunity to target these tissues passively and accumulate in them. In addition, it has been shown that stealth particles can penetrate the brain tissue of healthy animals, even though the blood–brain barrier (BBB) is reputed to be insurmountable for most therapeutic molecules [8]. The mechanism for this has not yet been perfectly understood, but these results open the way to promising diagnostic and therapeutic prospects.

11.2.3 Active Targeting by Coupling with Molecular Recognition Ligands

To target a given cell population, e.g., tumour cells, active targeting strategies must be developed on a case by case basis. The idea is to fix molecules on the particle surface that will bind specifically to receptors at the surface of the targeted cells, using some molecular recognition mechanism like the antigen–antibody interaction (see Fig. 11.2). It must be possible to graft these ligands onto the surface of the

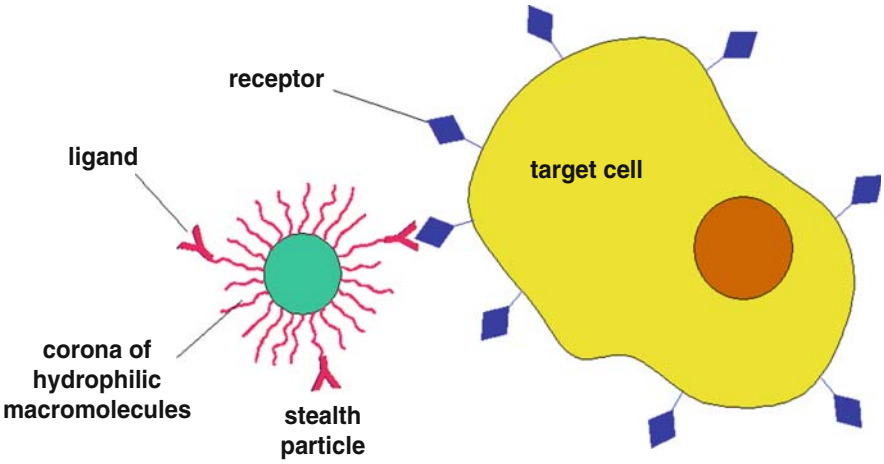

Fig. 11.2 Schematic view of molecular recognition applied to delivery of a nanoparticle to a target cell. The ligand grafted at the particle surface must be specific to the receptors present on the surface of the target cell

particles in such a way that they do not lose their addressing function, nor compromise the stealth of the system as a whole. This is why antibodies are still rarely used, being bulky (about 20 nm long), expensive to synthesise, and risky from the point of view of the immune response [9]. They also have the disadvantage of being very (or even too) specific to antigen epitopes, which for their part evolve in time. This is why peptides, sugars, or small molecules like folic acid are currently favoured. The latter is a group B vitamin, essential to the mechanism of cell division. Folic acid receptors are thus overexpressed at the surface of the cells most needing them, such as tumour cells [10]. This means that nanoparticles carrying folic acid at the surface should statistically end up more specifically on tumour cells.

11.2.4 Alternatives to Active Targeting

The challenge of active targeting has clearly not yet been met, and this for several reasons which depend in particular on the type of pathology treated. Current solutions are therefore more rudimentary and as a result display a certain number of drawbacks.

In cancerology, chemotherapy (alternative or complementary to radiotherapy and surgery) is administered systemically. The active principles are thus distributed throughout all the organs. They do of course treat pathological regions, but they also have side-effects in healthy organs, e.g., alopecia (hair loss), increased risk of infection, and digestive problems, among others.

Still at the experimental stage, the idea of physical targeting is to exploit the fact that tissues are transparent to magnetic fields and do not themselves contain a

magnetic component. One can thus envisage the possibility of dragging magnetic carriers, initially injected into the blood flow, toward target organs by creating magnetic field gradients, and thereby concentrating them in these organs for as long as the field is applied. This strategy has been under development for several decades. In most cases, the magnetic field gradient is produced by a strong permanent magnet (made from Nd–Fe–B) fixed to the outside of the body, close to the targeted organ. As an example, to concentrate carriers based on magnetite Fe_3O_4, it has been shown that the magnetic flux density must be of the order of 0.2 T in the target, with gradients of the order of 8 $\mathrm{T\,m^{-1}}$ for the femoral arteries and greater than 100 $\mathrm{T\,m^{-1}}$ for the carotid arteries [11]. The main difficulties remaining are lack of stealth with regard to the MPS, sporadic aggregation effects leading to a risk of embolus, the development of sufficiently powerful magnets for targeting to deep tissues, and somewhat disappointing results obtained as yet on large animals [12].

The last alternative is intratumoral injection, which is only applicable to solid and accessible tumours, and has the disadvantage that tumour cells may then migrate along the path taken by the needle.

11.2.5 Overview of Commercially Available and Forthcoming Formulations

Contrast Enhancement in MRI

Magnetic resonance imaging (MRI) exploits the properties of nuclear magnetic resonance of components of the human body, and in particular the protons in water contained in the tissues, lipid membranes, proteins, etc. [11]. It is based on the same idea as NMR spectroscopy used in chemical analysis, combining a strong static magnetic field B_0 (up to 2 T in standard hospital equipment) and a perpendicular radiofrequency field (5–100 MHz). After the radiofrequency pulse, the protons spins seek to realign with the static field B_0. This relaxation phenomenon can be decomposed into two independent phenomena: longitudinal relaxation which corresponds to a gradual increase in the longitudinal component of the magnetisation, and transverse relaxation which is the gradual decrease of the transverse component. They are characterised respectively by the T_1 relaxation time (the time required for 63% of the longitudinal component to be reacquired) and the T_2 relaxation time (the time required for 37% of the transverse component to disappear).

In order to reconstitute a 3D image, the NMR signals are collected in each elementary volume or voxel and correlated with the emission coordinates. To do this, B_0 or the radiofrequency field must vary in space so that each voxel has its own resonance frequency. The relaxation values are processed by a 2D Fourier transform. By playing on the sequence parameters, in particular the repetition time and delay time, the operator can obtain T_1- or T_2-weighted images.

Because they do not have the same relaxation times, tissues, fats, fluids, and so on can be distinguished from one another. However, as the MRI technique has been

developed, it has been noticed that exogenic contrast agents can significantly improve the appearance of tissue boundaries, and hence the reliability of the diagnosis. They act indirectly, because they are in fact substances whose magnetic properties can increase the relaxation rates of nearby protons. They are already used in some 40% of MRI examinations.

The first generation of these contrast agents comprises highly paramagnetic ions such as Gd^{3+}, with 7 unpaired electrons, stabilised in the form of molecular complexes in order to reduce their intrinsic toxicity. They have been used clinically since the end of the 1980s to mark extracellular and intravascular spaces (kidney function, state of the BBB, etc.). Their main disadvantages are a rather short plasma half-life (70–100 min), related to their low molar mass, and a relatively low contrast effect due to the fact that there is only one paramagnetic ion per molecule.

The next generation are the magnetic susceptibility contrast agents in the form of iron oxide nanoparticles (magnetite Fe_3O_4 and maghemite Fe_2O_3-γ), with diameters in the range 3–10 nm, these being clustered together to differing degrees and encapsulated within a hydrophilic dextran corona [13]. Since they are smaller than a magnetic domain, they lose their magnetisation as soon as the magnetic field is switched off. Their magnetic moment is nevertheless much higher than that of a paramagnetic compound, which is why they are known as superparamagnetic compounds. They are commonly called (ultrasmall) superparamagnetic iron oxides, or (U)SPIOs for short. They are obtained very simply by a one-step process, involving an alkaline coprecipitation of Fe(II) and Fe(III) precursors in an aqueous solution of dextran macromolecules [1]. The latter have the role of:

- limiting particle growth,
- stabilising the particles sterically,
- later, in vivo, improving biocompatibility and preventing opsonisation phenomena.

They are administered by perfusion at an average concentration of 1 mg of iron per kg of body mass. After endocytosis by the macrophages, they end up being metabolised in the lysosomes. So after solubilisation, the metal ions go into the iron pool of the organism, which is estimated at around 3 500 mg per person. Depending on the required diagnosis, one uses either SPIOs like Endorem (registered trade mark of Guerbet), or USPIOs like Sinerem (Guerbet). SPIOs have a hydrodynamic volume greater than 40 nm and, despite the presence of dextran, are rapidly accumulated in the MPS organs with a plasma half-life shorter than 10 min. They thus provide a way of imaging the liver. USPIOs have a smaller hydrodynamic volume and as a consequence their plasma half-life is longer than 2 hr, whence it is possible to image the blood vessels (angiography). The smallest particles can escape from the blood compartment via the interstitium and enter the lymph system. After drainage or capture by macrophages, they accumulate in the lymph nodes, which can thereby be imaged (lymphography).

Current developments aim to create contrast agents carrying ligands and hence able to target specific cell populations or tissues, so that administered doses can be

reduced, while at the same time enhancing the signal from the target (molecular imaging) [14].

Controlled Drug Release

The main goal for pharmacology is the transport of therapeutic molecules to specific targets, or drug carrying. However, the way in which these molecules leave the carrier to be able to carry out their task is also important. Pharmaceutical entities allowing controlled release of drugs belong to a new generation of medicines, the aim being to control the rate, duration, and even the place where the active substance is released. These are hollow or solid capsules of submicron dimensions where the active principle is either located in the middle or distributed throughout the matrix, respectively. Toxicity and renal clearance time are thus considerably reduced. The main release mechanisms are diffusion through the wall or matrix, or erosion/dissolution of the latter. The kinetics can be accelerated locally under the effects of a temperature or pH variation, or momentarily under the effects of radiation. The role of these new pharmaceutical forms is to maintain a sufficient concentration of the drug in circulation in the blood, and at the same time reduce the frequency of administration.

With regard to formulation, these carriers can take the form of micelles (< 50 nm) made from surfactants or block copolymers, liposomes (50–500 nm) made from phospholipid bilayers (and hence able to fuse with cell membranes), and biodegradable polymer capsules (10–1000 nm) [15]. However, at the present time, only a small number of liposome formulations are commercially available in rather specific therapeutic areas, e.g., AmBisome (registered trademark of Gilead Sciences, Inc., for transport of amphotericin B, a drug with antifungal properties).

Magnetically Induced Hyperthermia

It has been known for more than 5 000 years that heat can treat a great many illnesses, in particular cancers [16]. In modern oncology, hyperthermia is one of the four main therapeutic solutions, along with surgical ablation, radiotherapy, and chemotherapy, which are often combined. The most recent hyperthermic techniques fall into three categories:

- contact with a liquid or body heated from the outside,
- heating without contact by means of devices able to dispense energy remotely,
- implantation of optical fibres, antennas, or mediators in the human body, able in vivo to transport or convert energy, controlled from the outside, into heat.

These techniques are generally difficult to implement, some of them involving surgery. Furthermore, they are far from being totally and universally effective, i.e., they would not apply to every type of cancer, whatever its position, extent, and stage of development.

Over several decades then, the idea of using particles small enough to reach all tissues via the blood compartment has been slowly progressing. Naturally, scientists have turned their attention to magnetic nanoparticles, for several reasons:

- the human body is transparent to the magnetic field,
- it contains no magnetic material,
- under certain amplitude and frequency conditions, magnetic materials can release heat in an alternating magnetic field.

At the present time, clinical developments have reached phase II in the context of multiform glioblastoma and prostate carcinoma at the La Charité Hospital in Berlin [17]. This new approach will be discussed in more detail in Sect. 11.3.

Photoinduced Hyperthermia (Photothermal Therapy)

This idea, more recent, uses absorption properties associated with the phenomenon of surface plasmon resonance (SPR) affecting metal nanoparticles [18]. For example, gold nanoparticles very quickly ($\sim$ 1 ps) convert laser light into heat. In addition, their cross-section is up to 5 orders of magnitude greater than that of conventional molecular photoabsorbers, which means that much lower radiation energies could be used. Finally, exploiting both the SPR scattering and absorption phenomena, the nanoparticles could be simultaneously imaged, thus combining diagnosis and therapy in the same method. The first in vitro trials have been positive, and this method should be applicable to skin cancers. However, for most cancers, the light must be able to pass through a tissue thickness anywhere between a few mm and a few cm. For this reason, research is turning toward the design of nanoparticles that can absorb in the near infrared, especially in the window 650–900 nm, where certain tissues are transparent over several cm. Now, by modifying the form and chemical composition of the particles, the SPR absorption band can be shifted into this range and this is how systems are being developed with silica/gold core–shell particles or gold rods. In vivo studies on mice show that the heating produced by infrared absorption of silica/gold nanoparticles under the effects of a diode laser (808 nm, 4 W/cm^2, 3 min) is enough to destroy cancer cells, while preserving healthy tissue in the neighbourhood [19]. In this study, remission was total, prolonging the lives of the treated animals by at least three months.

It should nevertheless by noted that this photothermal technique involves focusing the laser on the nanoparticles, and this in turn requires prior knowledge of their biodistribution (or the assumption that the boundary of the tumour is known and that the particles lie within it), which is not the case with magnetically induced hyperthermia.

Toward a Multipurpose Platform

Far upstream of in vivo trials, much effort is being expended to combine several diagnostic or therapeutic features in a single device. Chemists are particularly interested in carriers whose biodistribution could be monitored in real time and which could release an active principle when the particle reaches its target. As an example, some work has involved association of gold colloids with a thermosensitive hydrogel polymer such as poly(N-isopropylacrylamide-co-acrylamide), which has the particularity of contracting suddenly above a certain temperature called the lowest critical solution temperature (LCST) (see Sect. 10.3.2 of Chap. 10) [20]. This mechanical retraction phenomenon leads to the release of the molecules originally emprisoned between the macromolecules. With this system, one would thus be able to control drug release by optical illumination of the relevant region. Progress nevertheless remains to be made, in particular to improve the biocompatibility of these thermosensitive macromolecules. Another example concerns mesoporous silica particles, in which some of the pores contain magnetic nanoparticles, while the other pores could carry active principles [21]. MRI contrast has proved adequate to allow localisation of the carriers, and release of active principles by diffusion could be thermally activated on demand by heating via an alternating magnetic field.

11.2.6 Relative Importance of Toxicity

Clearly, non-cytotoxicity is an essential criterion for any formulation to be used in diagnosis, because by default the patient is considered healthy. For a therapeutic application, the physician making the prescription is faced with classic dilemma between risk and benefit. An active principle is specifically toxic for sick cells, but also for healthy cells in the vast majority of cases. For this reason, the dose and duration of treatment are essential elements of the prescription, in order to limit side-effects. In this context, a certain level of carrier toxicity can be tolerated. However, whenever targeting becomes 100% efficient, toxicity will become a secondary drawback.

Finally, before developing diagnostic or therapeutic products on humans, biologists will need to understand a certain number of mechanisms and validate a certain number of concepts in vitro or on animal models. In this case, they can often accept more cytotoxic nanoparticles, such as cadmium chalcogenide quantum dots, which could clearly never go to clinical use.

11.3 Magnetically Induced Hyperthermia

11.3.1 Therapeutic Advantages of Heat

In cancerology, thermotherapy is a heat therapy that is complementary to the conventional treatments, viz., surgery, radiotherapy, and chemotherapy. It in fact includes three types of treatment [22]:

- Hyperthermia (between 40 and 43°C), stimulating the immune system of the patient for an anticancer response and generally associated with radiotherapy or chemotherapy [22]. The synergy here is based on the fact that hypoxic (oxygen deficient) cancer cells are more resistant to radiation than euoxic (normally oxygenated) cancer cells, but more sensitive to heat [23]. A temperature rise modifies the blood flow, the vascular permeability, the metabolism, and the oxygenation of tumour cells [24], and for certain types of tumour, these changes significantly influence the response to radiotherapy and chemotherapy. The combined use of radiotherapy and hyperthermia, known as thermoradiotherapy, generally leads to an increased success rate for a given radiation level, or, for given success rate, to a reduction in the required radiation dose [25].
- Total thermoablation (above 46°C), corresponding to necrosis of all cells without distinction.
- Selective thermoablation (between 43 and 46°C) of tumour cells, applying the principle that cancer cells are more sensitive to heat than healthy cells [26]. The first clinical trials to exploit this selectivity date back to the 1970s [27]. It would seem that, in this temperature range, the functions of certain structural proteins and enzymes are modified, thereby altering the cell growth and differentiation, to the point of causing apoptosis (programmed cell death) [28].

It should be noted that the temperature range required for selective thermoablation of cancer cells is very close to that required for total thermoablation, whence the need for very tight control of the intratumoral temperature, to within one degree. Furthermore, hyperthermia would appear to be beneficial for the treatment of illnesses other than cancers, such as syphilitic paralysis and AIDS [29].

11.3.2 Different Methods and Their Limitations

Depending on the nature of the heat source and the region to be treated, a distinction is made between contact thermotherapy (with an external hot body or liquid), thermotherapy without contact (by focused ultrasound, microwaves, or infrared radiation), and thermotherapy via an implanted heat source (introduced into the patient's body) [28, 30]. In the latter case, the heat sources are radiofrequency or microwave antennas, laser fibres, or mediators, i.e., materials converting electrical or magnetic energy into heat.

Systems currently used in thermotherapy are mainly based on focused ultrasound, e.g., Ablatherm Intensity Focused Ultrasound [31], or electromagnetic radiation. However, these systems cannot cause strong heating of deep tumours without damaging the surrounding healthy tissue. This is why methods of thermotherapy are being developed with internalised heat sources, using antennas or mediators.

Macroscopic mediators, usually beads or rods made from NiCu, FePt, or PdCo alloys, with sizes ranging from a few mm to a few cm, are directly inserted into the tumour. There are several drawbacks, e.g., the technique is invasive, some tumours are rather inaccessible, there is a risk of the mediators migrating elsewhere, and the temperature distribution is not uniform, among others [30].

Micrometric or nanometric mediators are micro- or nanoparticles that can be injected in the form of colloidal suspensions. If they heat up under the effect of an electric field (materials with high electrical conductivity), they are called capacitive mediators. If they heat up under the effect of a magnetic field (magnetic materials), they are called inductive mediators. The disadvantage with capacitive mediators lies in the fact that the human body may itself be heated by eddy currents owing to its intrinsic ionic conductivity ($0.6\ \Omega^{-1}\mathrm{cm}^{-1}$), causing uncontrolled and non-uniform heating. In this respect, inductive mediators would appear to be more promising, since there are no endogenous magnetic compounds.

11.3.3 Mechanisms for Induction Losses in Magnetic Materials

Magnetic particles can heat up via three mechanisms, depending on the amplitude H and frequency ν of the field, as well as the characteristics of the particles, e.g., intrinsic magnetic properties and surface chemistry, and characteristics of the medium in which the particles are dispersed. The dissipated heat is determined in a calorimeter (see Fig. 11.3), from the curve showing the time dependence of the temperature rise. To be precise, it is calculated in the form of an intensive quantity called the specific loss power (SLP) or specific absorption rate (SAR), determined at a given temperature and calculated from

$$\mathrm{SLP}_{\mathrm{magn}}^{37^\circ\mathrm{C}} = \frac{C}{x_{\mathrm{magn}}} \left.\frac{\mathrm{d}T}{\mathrm{d}t}\right|_{37^\circ\mathrm{C}}, \tag{11.1}$$

where C is the heat capacity of the medium (generally taken to be that of water) and x_{magn} is the mass fraction of the magnetic element. The physical unit of the specific loss power is then the watt per gram of the magnetic element.

Dissipation by Hysteresis Losses

For diameters above about 10 nm, magnetic particles comprise several magnetic domains (Weiss domains) with their own spontaneous magnetisation. In the presence

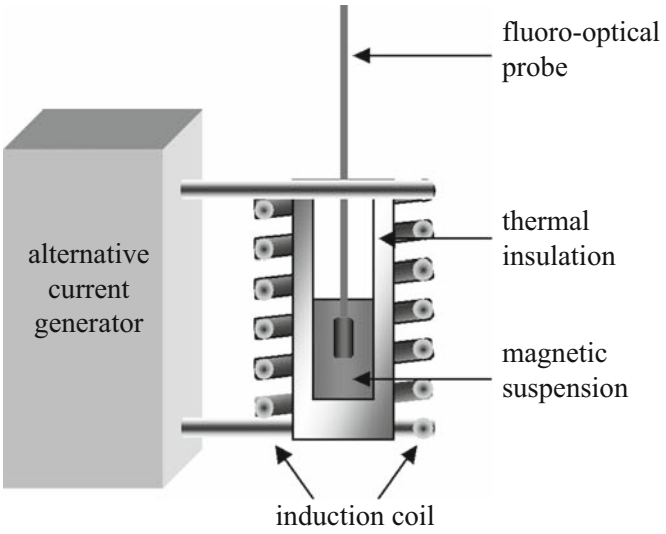

Fig. 11.3 Setup used to measure the temperature rise in samples subjected to an alternating magnetic field

of a static magnetic field of increasing strength, domains whose magnetisation has the same orientation as the magnetic field will grow in size, at the expense of the other domains, as the Bloch walls separating them gradually move out. By reversing the field, the phenomenon also reverses, and the graph showing the magnetisation of the material as a function of the field then takes the form of a hysteresis loop. This is the origin of the heat losses for an alternating magnetic field.

However, in order to correlate the magnetic characteristics of the particles with the resulting losses, the hysteresis cycle must be determined with a magnetic excitation of the same amplitude and frequency as the one used to generate the losses. The SLP is then proportional to the frequency and the area of the hysteresis loop [32].

Dissipation by Néel and Brownian Relaxation

For particle sizes below 10 nm, the particles are smaller than a magnetic domain. The area of the hysteresis cycle determined in a static field is zero and the particles are said to be superparamagnetic. They differ from the above ferromagnetic or ferrimagnetic particles because there is no remanent magnetisation once the field has been switched off. In an alternating magnetic field, the return to equilibrium, i.e., the disordered state, of the magnetic moments after perturbation occurs either by a motion of the particle (Brownian relaxation, also possible for multidomain particles), or by a motion of the magnetic moment (Néel relaxation).

Heating by Néel rotation occurs because, under the effect of an alternating magnetic field, the magnetisation of a magnetic domain, normally ‘frozen’ in one

direction, rotates with respect to the crystal lattice by overcoming the energy barrier E_a, the anisotropy energy. In the bulk model, $E_a = KV$, where K is the characteristic anisotropy constant of the material and V the volume of the magnetic core. This energy is dissipated in the form of heat when the magnetisation relaxes to its equilibrium orientation. The characteristic time for this relaxation, the Néel relaxation time τ_N, is given by

$$\tau_N = \tau_0 \exp \frac{KV}{kT} , \tag{11.2}$$

where the characteristic time τ_0 is 10^{-9} s and k is the Boltzmann constant [32]. The Néel frequency ν_N associated with this relaxation time, given by the relation $2\pi\nu_N\tau_N = 1$, corresponds to the frequency above which the heating by Néel rotation becomes dominant [33].

In the case of Brownian rotation, the magnetic particle rotates right round under the effect of the couple exerted by the external magnetic field on the particle magnetisation. In this case, the heat comes from friction between the particle and the surrounding medium. This rotation is characterised by the Brown relaxation time τ_B given by

$$\tau_B = \frac{3\eta V_h}{kT} , \tag{11.3}$$

where η is the viscosity of the surrounding medium and V_h the hydrodynamic volume of the particle [32]. This phenomenon is very sensitive to the viscosity of the medium, but also to the presence of molecules or macromolecules on the particle surface, and whether the particle is bound to the target or not. The Brown frequency ν_B, given by the relation $2\pi\nu_B\tau_B = 1$, is the frequency above which heating by Brown rotation becomes dominant [33].

Since these two processes act in parallel, the overall relaxation time is given by

$$\frac{1}{\tau} = \frac{1}{\tau_N} + \frac{1}{\tau_B} , \tag{11.4}$$

and since τ_N depends exponentially on the volume, while τ_B is directly proportional to it, one of the two terms will be negligible compared with the other and the relaxation dynamics of the particle will be governed by just one of these two phenomena.

11.3.4 Comparing Theory and Experiment

Up to the present time, work on magnetically induced hyperthermia has fallen into three categories [32]:

- In vivo studies by medical teams, aiming to show directly that this type of treatment can be effective.
- Synthesis of colloidal dispersions with optimal heating properties.
- Physical models for describing the mechanisms and determining the parameters governing heating.

Unfortunately, these three research communities often work in parallel, and do not move forward at the same rate. In particular, attempts to correlate theory with experiment are scarce and often not conclusive. However, the data needed to specify requirements for optimising the SLP are now available as a result of a certain number of increasingly accurate theoretical studies [32, 34–38].

In particular, in most situations, the SLP turns out to be proportional to the squared amplitude of the magnetic field. Depending on the order of magnitude of the relaxation time τ, losses are expected to increase linearly with the frequency and then level off when $2\pi\nu \gg 1/\tau$. A high anisotropy constant penalises Néel relaxation to the benefit of Brownian relaxation, whereas a high viscosity has the opposite effect. Finally, the volume, and hence the diameter, of the particles has a considerable influence on τ, and this is not without consequences for the heating of real colloidal dispersions, in which there is inevitably a distribution of particle sizes. Indeed, in particular for particles undergoing Néel relaxation, it is clear that only those particles whose volume corresponds to the applied frequency will contribute to heating, whence the need to prepare particle dispersions with as narrow a size distribution as possible [39].

For these reasons, future work concerns not only the synthesis of nanoparticles with controlled granulometry, surface chemistry, and colloidal stability, but also the development of equipment able to scan frequencies and validate optimal field conditions.

11.3.5 Physiological Constraints

Amplitude and Frequency of the Applied Field

In order to produce an alternating magnetic field, one must favour the magnetic component of the radiation over the electric component by using a solenoid (or air gap). From the physiological standpoint, applicable frequencies must be above 50 kHz to avoid neuromuscular electrostimulation effects, but less than 100 MHz to enter deeply enough into the body [40]. Moreover, for a coil of diameter about 30 cm, the product $H\nu$ must be less than 4.85×10^8 A/m s to allow one hour sessions without discomfort for the patient (otherwise non-uniform heating effects become uncomfortable) [41]. This product can be made an order of magnitude greater by using a coil of lesser diameter [42]. Indeed, the electrical component of the radiation cannot be completely eliminated, and there is a risk of generalised and non-unifom heating of the body tissues and fluids by eddy currents.

For these reasons, multidomain particles, which heat up via the dissipation mechanisms provided by hysteresis losses, or even Brownian relaxation, are in principle less favourable, requiring a large field amplitude (to produce the largest possible hysteresis cycle), and also a high frequency. Under physiological conditions, their SLP is therefore rather disappointing. On the other hand, superparamagnetic particles have the advantage of being efficient from 100 kHz.

In addition, it does seem somewhat risky to depend only on Brownian relaxation to reach sufficiently high values of the SLP from the therapeutic point of view, because unfortunately, once in the target tissues, or target cells, it is not obvious that the particles will be able to turn round owing to a high local viscosity or strong interactions with the cell membranes.

Controlling the Maximal Temperature in Vivo

The value of the specific loss power is a fundamental parameter for thermotherapeutic applications of magnetic particles. Indeed, the higher the SLP, the lower the dose of nanoparticles that will need to be injected into the patient. Therefore most current research aims primarily to increase the value of the SLP. However, there is an adverse effect in this venture: particles with high SLP heat very quickly, and it is not obvious that the temperature will remain within the thermotherapeutic range, without rising into the total thermoablation range.

A first solution might be to continuously monitor the temperature distribution and control it by adjusting the magnetic field. Unfortunately, thermometry is invasive, especially as the temperature must be taken at several different points. Through the temperature dependence of the proton relaxation time (or the chemical shift), it should be possible to use MRI to carry out continuous temperature monitoring, with instantaneous feedback on the power of the magnetic field. This method is currently being developed on some thermotherapy systems using focused ultrasound, but in the situation of magnetically induced hyperthermia, it seems likely that the nanoparticles themselves would perturb the images and hence thwart accurate measurement [43].

Another solution would be to exploit the temperature dependence of the magnetic properties of some compounds. The idea would be to use ferromagnetic or ferrimagnetic nanoparticles with a Curie temperature adjusted to just above the temperature that should not be exceeded in vivo. In this way, should that temperature be reached, each nanoparticle mediator would function as its own fuse by losing its magnetic properties and hence also its heating capability [44]. Studies are currently being undertaken on manganese perovskites (see Fig. 11.4) [45, 46].

Controlling the Temperature Distribution in Vivo

The temperature distribution in vivo depends strongly on the nanoparticle distribution, the heat capacity of the tissues, their conductivity, their shape, and the way they are irrigated. Calculations can simulate the spatial distribution of the temperature field within poorly perfused muscle tissues [47]. In addition, it has been shown that, during a hyperthermia session, the presence of large vessels entering the heated region can constitute a significant source of non-uniformity in the temperature (convection cooling) and a risk of underdose [48, 49]. In this case, one solution might be to heat the blood upstream of the tumour. In general, well-perfused

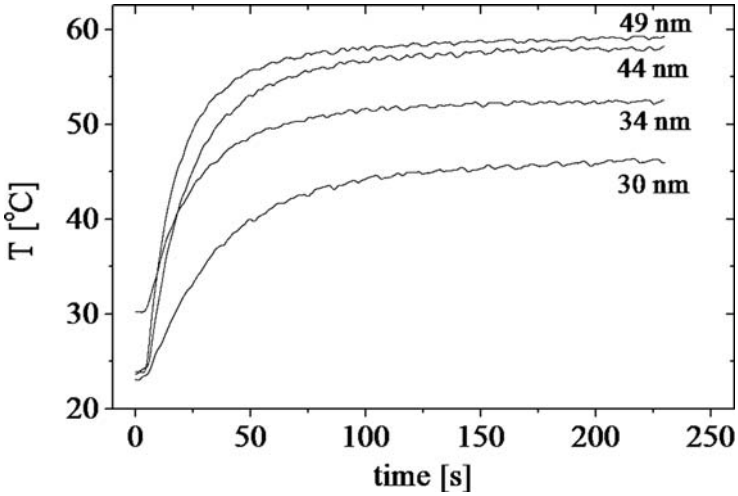

Fig. 11.4 Evolution of the temperature of a colloidal dispersion of particles of $La_{0.75}Sr_{0.25}MnO_3$ with different diameters ([Mn] = 2.4 $g_{Mn}L^{-1}$, H_{max} = 70 kA/m and ν = 108 kHz) [45]. The maximal temperature reached depends on the size of the crystallites

tumours are more likely to be heated uniformly. Other work has shown that a shorter heating period (1–10 s) at a higher temperature is more effective than slow heating for obtaining a uniform temperature [50]. Recent work in vitro has also focused on the heating properties of maghemite in living cells [50].

11.3.6 Some Formulations Under Development or Undergoing Clinical Assessment

The first in vivo thermoablation experiment dates back to 1957 [52]. Today, in oncology, there are at least two complementary strategies being developed by Australian and German research teams, using nanocomposite magnetic microparticles and surface-modified superparamagnetic nanoparticles, respectively.

Nanocomposite Magnetic Microparticles Administered by Arterial Embolisation

This form of administration exploits the fact that liver tumours are irrigated mainly by the hepatic arterial system, while healthy liver cells are supplied by the portal vein system [53]. Experiments have been carried out on rabbits [54] and pigs [55], yielding iron concentrations in the tumour five times higher than in the healthy tissues. The first experiments used submicron maghemite particles (diameter 150 nm)

suspended in lipiodol (a mixture of iodine and vegetable oil, known for its ability to trap anticancer molecules in the chaotic network of blood vessels in tumours) [56]. Under magnetic induction conditions of 53 kHz and 30 kA/m, an intratumoral temperature of 48°C was reached in 5 minutes. However, since lipiodol proved to be too vaso-occlusive for hepatic use, leading to extended necrosis [57], the maghemite particles were then encapsulated in polymer beads of average diameter 32 μm, dispersed in a 1% aqueous solution of dispersant (Tween) (Thermospheres, registered trademark of SIRTeX Medical Ltd, Sidney). These beads have been shown to be non-toxic and well tolerated. Following these positive results on animals, the company SIRTeX Medical Ltd is now investigating the potential for this technology on humans, and the possibility of extending its application to cancers other than liver cancer.

Superparamagnetic Nanoparticles Injected Directly into the Tumour

The first studies used magnetite nanoparticles surrounded by a dextran corona, similar to the (U)SPIO contrast agents developed for MRI [58]. In vitro experiments showed that these nanoparticles were internalised by cancer cells. Applying an alternating magnetic field (520 kHz, 7–13 kA/m) caused the tumour to regress in the same way as it would in a bath of hot water. Subsequently, dextran was abandoned because it would appear to degrade too soon.

It was then shown that direct injection of particles into the tumour had an advantage (still not fully understood) called the thermal bystander effect: even though the particles are concentrated at the deposition points before applying the alternating magnetic field, they distribute themselves much more uniformly after the first application of the alternating magnetic field [22, 59].

Recently, using magnetite particles coated with aminosilanes (magnetic core 10 nm, hydrodynamic diameter 30 nm), whose surface is thus positively charged under physiological conditions, promising results have been obtained in vitro [60, 61] and in animals [62, 63], while the first clinical trials have proved encouraging on several types of cancer [64–66]. Since then, a company called MagForce Nanotechnologies AG (Berlin) [17] was created and the first human-scale magnetic field applicator MFH® 300F put together [67]. To solve the problem of the temperature distribution, the inventors used numerical simulations based on tomographic images of the tumour which allow them to optimise nanoparticle deposition (see Fig. 11.5) [68].

11.4 Short and Mid-Term Prospects

It is clear that, if magnetically induced hyperthermia is to achieve the desired results, a certain number of aspects need to be improved. This includes the synthesis and properties of the mediators, but also our understanding of the physics involved in

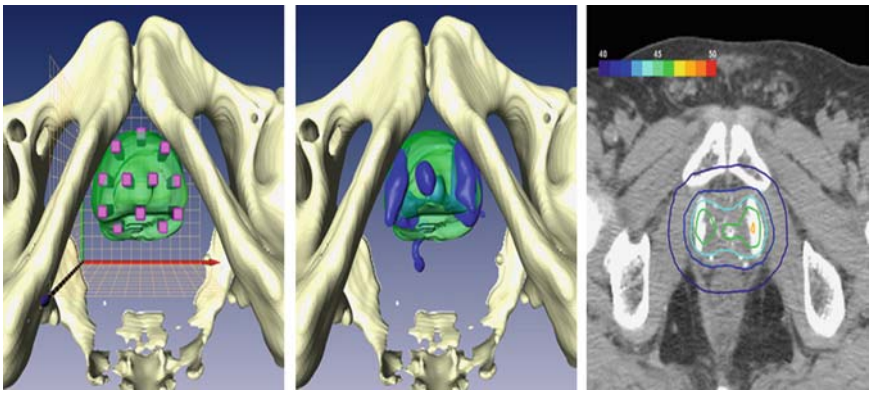

Fig. 11.5 *Left*: Tomography of a prostate (*green region*), indicating the path of needles used to deposit 0.5–1 mL (per path) of the particle dispersion. *Center*: Tomography after injecting the nanoparticles (*blue*). *Right*: Image showing the nanoparticle deposits (*white* because denser than the prostate tissue) and calculated isotherms. From [68]. Copyright Elsevier

the dissipation phenomena, modelling of the in vivo temperature distribution, and development of a safe, efficient, and reasonably priced active targeting strategy.

11.4.1 Mediators

Chemists now know that, apart from the criteria of average granulometry and colloidal stability, the following conditions must necessarily be fulfilled:

- The narrowest size distribution possible, in order to specify the best magnetic field characteristics for optimising the SLP.
- More 'intelligent' magnetic properties, so that the Curie temperature can be adjusted and the appropriate therapeutic temperature can be self-regulated.
- Surface functionality for grafting on hydrophilic macromolecules to improve stealth and ligands to implement active targeting.
- The design of multipurpose platforms that can combine thermotherapeutic treatment with chemotherapeutic action, using local heat production to release active principles.

Some of these constraints pull in opposite directions, e.g., maximal SLP values are attained for superparamagnetic compounds, while in principle only ferromagnetic and ferrimagnetic compounds allow self-regulation. The aim here will thus be to find the best possible compromise.

11.4.2 Physics of Magnetic Dissipation Phenomena and Models for the in Vivo Temperature Distribution

If the heating capacities of the mediators are to be improved, the models used up to now will need to be refined and extended, to increase their predictive power and to cover the whole range of sizes and magnetic behaviour. They must be systematically validated by experiment and the difficulties inherent in this approach will not be overcome without close collaboration between chemists and physicists, both in theory and practice [32].

Modelling the temperature distribution in complex systems remains a priority for determining the optimal ways of depositing energy, e.g., continuously or in pulses.

11.4.3 Targeting Strategies

There is a genuine debate over the best means of administration [32]. Local injection can deposit the particles in a single place and control the amount, but it may allow cancer cells to disseminate over the path taken by the needle. Intravenous injection is non-invasive and provides a way of reaching the whole organism. However, a large fraction of the product is transported to the liver and spleen and the tumour dose is therefore difficult to control.

Active targeting would certainly be the ideal solution, but unfortunately it has not yet proven itself in vivo, despite several preliminary studies in vitro [69, 70]. There is little doubt that mixed strategies should be developed, combining active targeting and physical targeting (concentrating mediators in the relevant region by means of an external magnet).

Note also that, with regard to active targeting of a given cell population, one question remains unanswered: is intracellular thermotherapy, i.e., obtained with magnetic particles internalised within cells, more efficient than extracellular thermotherapy? According to a theoretical model, it would seem that there is no thermotherapeutic effect on the nanoscale (the scale of the particles) or the microscale (the scale of the cells). In fact, thermotherapy operates on the millimeter scale (the scale of the tumour), since the thermal insulation behaviour of the cell membrane is negligible [71]. Any difference that may be observed between intracellular and extracellular thermotherapy would thus appear to be due to a mechanical effect of rotation or vibration of the particles in the cell.

11.4.4 System Applying the Alternating Magnetic Field

The dimensions of the human body raise the question of what kind of equipment should be used to produce the magnetic field [32]. A local device has the advantage

of providing better control over the applied field, but it may involve an invasive act of surgery, while application to the body as a whole is simpler and non-invasive, but requires higher field amplitudes, so that one must ensure that the treatment is harmless for the rest of the organism.

References

1. Duguet, E., Treguer-Delapierre, M., Delville, M.H.: In: *Les Nanosciences*. Vol. 3. *Nanobiotechnologies et Nanobiologie*, ed. by M. Lahmani, P. Boisseau, P. Houdy P. (Belin, Paris, 2007) pp. 145. To be published in English by Springer, Heidelberg (2009)
2. Mornet, S., Vasseur, S., Grasset, F., Duguet, E.: J. Mater. Chem. **14**, 2161 (2004)
3. Duguet, E., Mornet, S., Vasseur, S., Devoisselle, J.M.: Nanomed. **1**, 157 (2006)
4. Freitas, R.A.: *Nanomedicine*, Vol. I, *Basic Capabilities* (Landes Bioscience, Georgetown, 1999)
5. Monfardini, C., Veroneses, F.M.: Bioconj. Chem. **9**, 418 (1998)
6. Hume, D.A., Ross, I.L., Himes, S.R., Sasmono, R.T., Wells, C.A., Ravasi, T.: J. Leukocyte Biol. **72**, 621 (2002)
7. Moghimi, S.M., Hunter, A.C., Murray, J.C.: Pharmacol. Rev. **53**, 283 (2001)
8. Kreuter, J.: Adv. Drug Delivery Rev. **47**, 65 (2001)
9. Weinmann, H.J., Ebert, W., Misselwitz, B., Schmitt-Willich, H.: Eur. J. Radiol. **46**, 33 (2003)
10. Stella, B., Arpicco, S., Peracchia, M.T., Desmaëlle, D., Hoebeke, J., Renoir, M., D'Angelo, J., Cattel, L., Couvreur, P.: J. Pharm. Sci. **89**, 1452 (2000)
11. Pankhurst, Q.A., Connolly, J., Jones, S.K., Dobson, J.: J. Phys. D Appl. Phys. **36**, R167 (2003)
12. Sonvico, F., Dubernet, C., Colombo, P., Couvreur, P.: Current Pharmaceutical Design **11**, 2091 (2005)
13. Bonnemain, B.: J. Drug Targeting **6**, 167 (1998)
14. Corot, C., Robert, P., Idée, J.M., Port, M.: Adv. Drug Delivery Rev. **58**, 1471 (2006)
15. Couvreur, P., Brigger, I., Dubernet, C.: Adv. Drug Delivery Rev. **54**, 631 (2002)
16. Andrä, W.: In: *Magnetism in Medicine*, ed. by W. Andrä, H. Nowak (Wiley-VCH, 1998) p. 455
17. `http://www.magforce.de/` consulted on 11 February 2009
18. Jain, P.K., El-Sayed, I.H., El-Sayed, M.A.: Nanotoday **2**, 18 (2007)
19. O'Neal, P., Hirsch, L.R., Halas, N.J., Payne, J.D., West, J.L.: Cancer Lett. **209**, 171 (2004)
20. Sershen, S.R., Westcott, S.L., Halas, N.J., West, J.L.: J. Biomed. Mater. Res. **51**, 293 (2000)
21. Julián-López, B., Boissière, C., Grosso, D., Chanéac, C., Vasseur, S., Duguet, E., Sanchez, C.: J. Mater. Chem. **17**, 1563 (2007)
22. Jordan, A., Scholz, R., Wust, P., Fähling, H., Felix, R.: J. Magn. Magn. Mater. **201**, 413–419 (1999)
23. Hofer, K.G.: *Abstracts of the 4th International Conference on Scientific and Clinical Applications of Magnetic Carriers* 78 (2002)
24. Vujaskovic, Z., Song, C.W., Roti, J.L.: Int. J. Hyperthermia **20**, 163 (2004)
25. Dubois, J.B.: Bull. Cancer/Radiother. **82**, 207 (1995)
26. Bacri, J.C., Da Silva, M.F., Perzynski, R., Pons, J.N., Roger, J., Sabolovic, D., Halbreich, A.: In: *Scientific and Clinical Applications of Magnetic Carriers*, ed. by Häfeli et al. (Plenum Press, New York, 1997) p. 597
27. Overgaard, K., Overgaard, J.: Europ. J. Cancer **8**, 65 (1972)
28. Moroz, P., Jones, S.K., Gray, B.N.: J. Surg. Oncol. **77**, 259 (2001)
29. Owens, S.D., Gasper, P.W.: Med. Hypotheses **44**, 235 (1995)
30. Moroz, P., Jones, S.K., Gray, B.N.: Int. J. Hyperthermia **18**, 267 (2002)
31. `http://www.hifu-planet.com` consulted on 11 February 2009

32. Fortin, J.P.: PhD Thesis at University Paris 7 Denis Diderot (2007)
33. Fannin, P.C., Charles, S.W.: J. Phys. D Appl. Phys. **24**, 76 (1991)
34. Jordan, A., Wust, P., Fähling, H., John, W., Hinz, A., Felix, R.: Int. J. Hyperthermia **9**, 51 (1993)
35. Hergt, R., Andrä, W., d'Ambly, C.G., Hilger, I., Kaiser, W.A., Richter, U., Schmidt, H.G.: IEEE Trans. Magn. **34**, 3745 (1998)
36. Rosensweig, R.E.: Magnetohydrodynamics **36**, 303 (2000)
37. Rosensweig, R.E.: J. Magn. Magn. Mater. **252**, 370 (2002)
38. Raikher, Y.L., Stepanov, V.I., Perzynski, R.: Physica B **343**, 262 (2004)
39. Jordan, A., Rheinländer, T., Waldöfner, N., Scholz, R.: J. Nanoparticle Res. **5**, 597 (2003)
40. Hill, D.A.: Bioelectromagnetics **6**, 33 (1985)
41. Brezovich, I.A.: Medical Physics Monograph **16**, 82 (1988)
42. Hergt, R., Dutz, S.: J. Magn. Magn. Mater. **311**, 187 (2007)
43. Quesson, B., Vimeux, F., Salomir, R., de Zwart, J.A., Moonen, C.T.W.: Magn. Reson. Med. **47**, 1065 (2002)
44. Kuznetsov, A.A., Shlyakhtin, O.A., Brusentsov, N.A., Kuznetsov, O.A.: Eur. Cells Mater. **3**, 75 (2002)
45. Pollert, E., Knížek, K., Maryško, M., Kašpar, P., Vasseur, S., Duguet, E.: J. Magn. Magn. Mater. **316**, 122 (2007)
46. Prasad, N.K., Rathinasamy, K., Panda, D., Bahadur, D.: J. Biomed. Mater. Res. B **85**, 409 (2008)
47. Andrä, W., d'Ambly, C.G., Hergt, R., Hilger, I., Kaiser, W.A.: J. Magn. Magn. Mater. **194**, 197 (1999)
48. Roemer, R.B.: Int. J. Hyperthermia **7**, 317 (1991)
49. Crezee, J.: Phys. Med. Biol. **37**, 1321 (1992)
50. Hunt, J.W., Lalonde, R., Ginsberg, H., Urchuk, S., Worthington, A.: Int. J. Hyperthermia **7**, 703 (1991)
51. Fortin, J.P., Gazeau, F., Wilhem, C.: Eur. Biophys J. **37**, 223 (2008)
52. Gilchrist, R.K., Medal, R., Shorey, W.D., Hanselman, R.C., Parrott, J.C., Taylor, C.B.: Ann. Surg. **146**, 596 (1957)
53. Moroz, P., Jones, S.K., Winter, J., Gray, B.N.: J. Surg. Oncol. **78**, 22 (2001)
54. Moroz, P., Jones, S.K., Gray, B.N.: J. Surg. Oncol. **80**, 149 (2002)
55. Moroz, P., Jones, S.K., Gray, B.N.: J. Surg. Res. **105**, 209 (2002)
56. Moroz, P., Pardoe, H., Jones, S.K., St. Pierre, T.G., Song, S., Gray, B.N.: Phys. Med. Biol. **47**, 1591 (2002)
57. Moroz, P., Jones, S.K., Metcalf, C., Gray, B.N.: Int. J. Hyperthermia **19**, 23 (2003)
58. Gordon, R.T., Hines, J.R., Gordon, D.: Med. Hypotheses **5**, 83 (1979)
59. Jordan, A., Scholz, R., Wust, P., Fähling, H., Krause, J., Wlodarczyk, W., Sander, B., Vogl, T., Felix, R.: Int. J. Hyperthermia **13**, 587 (1997)
60. Jordan, A., Scholz, R., Wust, P., Schirra, H., Schiestel, T., Schmidt, H., Felix, R.: J. Magn. Magn. Mater. **194**, 185 (1999)
61. Jordan, A., Wust, P., Scholz, R., Tesche, B., Fahling, H., Mitrovics, T., Vogl, T., Cervos-Navarro, J., Felix, R.: Int. J. Hyperthermia **12**, 705 (1996)
62. Johannsen, M., Thiesen, B., Jordan, A., Taymoorian, K., Gneveckow, U., Waldöfner, N., Scholz, R., Koch, M., Lein, M., Jung, K., Loening, S.: Prostate **64**, 283 (2005)
63. Johannsen, M., Thiesen, B., Gneveckow, U., Taymoorian, K., Waldöfner, N., Scholz, R., Deger, S., Jung, K., Loening, S.A., Jordan, A.: Prostate **66**, 97 (2006)
64. Gneveckow, U., Jordan, A., Scholz, R., Brüss, V., Waldöfner, N., Ricke, J., Feussner, A., Hildebrandt, B., Rau, B., Wust, P.: Med. Phys. **31**, 1444 (2004)
65. Johannsen, M., Gneveckow, U., Eckelt, L., Feussner, A., Waldöfner, N., Scholz, R., Deger, S., Wust, P., Loening, S.A., Jordan, A.: Int. J. Hyperthermia **21**, 637 (2005)
66. Wust, P., Gneveckow, U., Johannsen, M., Böhmer, D., Henkel, T., Kahmann,F., Sehouli, J., Felix, R., Ricke, J., Jordan, A.: Int. J. Hyperthermia **22**, 673 (2006)

67. Jordan, A., Scholz, R., Maier-Hauff, K., Johannsen, M., Wust, P., Nadobny, J., Schirra, H., Schmidt, H., Deger, S., Loening, S., Lanksch, W., Felix, R.: J. Magn. Magn. Mater. **225**, 118 (2001)
68. Johannsen, M., Gneveckow, U., Thiesen, B., Taymoorian, K., Cho, C.H., Waldöfner, N., Scholz, R., Jordan, A., Loening, S.A., Wust, P.: Eur. Urology **52**, 1653 (2007)
69. Suzuki, M., Shinkai, M., Kamihira, M., Kobayashi, T.: Biotechnol. Appl. Biochem. **21**, 335 (1995)
70. Sonvico, F., Mornet, S., Vasseur, S., Dubernet, C., Jaillard, D., Degrouard, J., Hoebeke, J., Duguet, E., Colombo, P., Couvreur, P.: Bioconj. Chem. **16**, 1181 (2005)
71. Rabin, Y.: Int. J. Hyperthermia **18**, 194 (2002)

Chapter 12
Accounting for Heat Transfer Problems in the Semiconductor Industry

Christian Brylinski

12.1 Introduction

Electronics has become omnipresent in our everyday lives. Occurring in all modern machines in the form of systems, functions, and components, it is gradually supplementing or replacing those functions previously carried out exclusively by mechanics, electromechanics, hydraulics, and pneumatics, by making the processes faster, more flexible, and safer in a quite spectacular way, and enriching the interaction between human and machine, until it has become a key feature of innovation and competitivity in all sectors of the economy.

The preliminary 'electronification' of existing systems is quickly followed by ever more sophisticated attempts to integrate electronic components and functions as close as possible to the target information sources and the devices to be operated, positioning the information processing and storage centers (processor and memory) as judiciously as possible. In this way, all kinds of chip are taken away from the sheltered conditions of specialised containers and end up having to operate in whatever environment prevails at the heart of the system they are designed to serve. In high speed trains, the encapsulated chips of the power switches are in contact with the alternator, at temperatures that sometimes reach 300°C, while those controlling car ignition must resist humidity and corrosion, and the power transistors in radars and lasers of on-board lidar systems have to operate at high altitudes, at sea, or in the field.

As microelectronics has been developed, chips have been specialised according to function:

- Some store and process information. Today, these are VLSI (very large scale integration) integrated circuits, made from silicon and using CMOS (complementary metal oxide semiconductor) technology.
- Others, generally using III–V semiconductors, emit light via light-emitting diodes or LEDs, or their variant, laser diodes. The former are at present assailing the lighting market with the long term objective of achieving domination. The latter,

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 367
DOI 10.1007/978-3-642-04258-4_12, © Springer-Verlag Berlin Heidelberg 2009

which already lead the field for mass data storage on disk, are preparing a major offensive in the area of image projection in the coming decade.
- For other chips, with rather similar structures and made from similar materials, the same III–V semiconductors are supplementing or replacing silicon in the conversion of optical data and energy into their electrical analogues.
- Some large chips, with diameters up to 100 mm for a single chip, deal with large amounts of energy, up to several megajoules, to control and convert the distributed electrical energy. To carry out these functions, silicon is not the ideal semiconductor. Alternative semiconducting materials with wide band gap are under investigation.
- At ultrahigh frequencies, e.g., in base stations for cell phones, when switching cannot keep up with the required rate, the energy is modulated analogically in chips specially optimised to obtain maximal linearity with the best possible energy efficiency. For an increasing number of applications, this modulation is achieved by transistors. Depending on the power and frequency, silicon, gallium arsenide, and indium phosphide are favoured.
- Finally, for the highest powers and the highest frequencies, there is no choice in 2008 but to implement electron beams in vacuum inside magnetrons, klystrons, gyrotrons, and travelling wave tubes. This is the case for magnetrons in microwave ovens, which typically deliver 1 kW @2.45 GHz @10 EUR, for travelling wave components in telecommunications satellites which operate at powers up to 500 W @30 GHz @10 kEUR, and also for synchrotrons which require around a megawatt at 350 MHz, and even for fusion research, where a single gyrotron supplies a megawatt at 140 GHz to contribute to heating the plasma.

Each kind of chip or device can only work in an optimal and long-lasting way if it has an environment that is compatible with its survival and 'metabolic' operation. Temperature is one of the key features of this environment. In the following, we shall review the main types of chip available or under development in 2008, presenting the specific requirements of each with regard to its thermal environment, together with known solutions and predictable developments for maintaining this environment under the technical and economic conditions corresponding to each context.

However, before going into more detail, it is useful to point out several general trends that currently define the task of those who design electronic systems, functions, components, and chips.

12.2 General Trends

12.2.1 Miniaturisation

An obvious trend is miniaturisation. Its rate and extent depend on the possibilities offered by the relevant physics of the functions that need to be carried out. For memory and processing chips, no serious obstacle has yet resisted human inventiveness.

For ultrahigh frequencies and optics, the wavelength constitutes an a priori datum, and considerable effort is required to get round it, often beyond the means available in this sector, where material and component markets are very limited, even on a worldwide scale.

Miniaturisation is always accompanied by an increase in the density of elementary functions, but also, as often as possible, by a decrease in the power handled and dissipated within each elementary cell. However, in the long term, it is observed that the rate of growth of the density exceeds the rate of fall in the local dissipation, so that there is an unavoidable increase in the density of heat dissipated in the bulk and per unit surface area in objects produced by the electronics industry.

12.2.2 Rising Frequencies

A second trend, related to the first, is the rising frequency of operations carried out by chips. In processors, it is of course motivated by the insatiable need to boost computation power. In telecommunications, an important factor is antenna size, proportional to the wavelength, while another constraint comes from overcrowding of the electromagnetic spectrum in free space, pushing new users to ever higher atmospheric levels. For example, cell phone technology, which began its prehistory prior to 1990 below 500 MHz, placed its GSM system around 900 MHz, then the UMTS network around 1.8 and 2.2 GHz, and is currently entering the 3–4 GHz bands for Wimax systems. Concerning professional systems, base stations communicate with each other at around 40 GHz, and with satellites between 8 and 30 GHz, and we are currently beginning to exploit the discrete 60 GHz band for local professional networks on air. A frequency of 77 GHz has been chosen in Europe for the anti-collision radars on vehicles, and some millimeter wave military systems operate around 94 GHz, as we await the predictable exploitation of the ultimate windows around 140 and 220 GHz.

With regard to the control of electrical energy, rising frequencies allow a considerable reduction in the weight and volume of passive elements. For example, for the same power, voltage, and current, a ferrite transformer operating around 30 kHz occupies a volume at least 10 times smaller and has a mass at least 10 times less than an iron alloy magnetic circuit transformer operating at 50 Hz.

Ultimately, for a given semiconductor material operating at a given voltage and current in a given environment, an elementary operation corresponds to the dissipation of a finite minimal energy with an incompressible lower threshold. It follows that, in order to carry out the same function, heat dissipation is higher at high frequency, for the same power. The gain in volume and mass on passive components is obtained at the cost of a further constraint on active semiconductor components.

12.2.3 Heterogeneous Integration

A third general trend is the integration of dissimilar functions within the same object. The integration of light sensors in data processing and storage chips underlies digital photography. The integration on silicon of all kinds of passive elements, e.g., resistances, capacitors, inductances, transformers, protection diodes, etc., has already made it possible to eliminate 90% of the component assembly operations in portable telephones. In spite of the formidable resistance of physics to these objectives over the past 20 years, the integration of light-emission functions within silicon chips is still the subject of intense research and industrial projects. So-called 'intelligent' electromechanical entities have been made, associating data processing functions with systems for acting on liquids, gases and solids. These are known as microelectromechanical systems, or MEMS, and their recently designed nanoscale prototypes, the NEMS. Finally, we must not forget the huge potential extension of the field of application of semiconductors represented by biological microsystems and interfaces between electronics and the living world. Here, of course, the survival of this living material is only possible within a very precise temperature range. Heterogeneous integration is always driven by a concern for economy. In general, it tends to downgrade the properties of each of the integrated functions, mainly because the optimal environments for each of the functional semiconducting materials are different, and the common environment imposed by such promiscuity can only be an imperfect compromise.

On one of the shared scales, the one regarding heat transfer, these three general trends of miniaturisation, rising frequency, and heterogeneous integration tend to combine for the worst, and this connivance suggests that very serious difficulties lie in wait for future generations of designers. In 2008, a top-of-the-range personal computer already includes fans and caloducts. Clearly, nobody really knows today what the computer of 2020 will need to survive the predictably enormous excess of calories it will generate.

12.3 Heat Transport

12.3.1 Heat Conduction

The materials used in the active part of an electronic or optoelectronic component are insulators or semiconductors. Metallic materials may be in contact with or in close proximity to the semiconductor, but the relevant physical effects occur in the semiconductor or at the interface between metal and semiconductor, in the case of Schottky contact components, e.g., Schottky rectifiers and MESFET transistors.

In semiconductors and insulators, heat conduction occurs almost exclusively through interatomic vibrations. Conduction by electrons only plays a significant

role in metals, and it only dominates in metals with very high electrical conductivity, such as copper, gold, silver, and aluminium, which are all used as on-chip electrical and/or thermal conductors, or in the various levels of interconnection in the architecture of electronic systems.

In the temperature range for operation of electronic components, typically between 200 and 500 K, the thermal conductivity of most standard insulators and semiconductors falls off quickly with the temperature according to a $T^{-\alpha}$ dependence, where α lies in the range 1–2 depending on the material and the temperature. The heating of semiconductor components is therefore self-impeding with regard to heat conduction, and this tendency is a possible cause of self-destruction.

The thermal conductivity at 300 K in standard monocrystalline semiconductors generally lies in the range 50–500 W/K m (150 W/K m for conventional silicon), with a considerable increase at low temperatures. Regarding the most widely used amorphous insulators in microelectronics and optoelectronics, values are closer to a few W/K m (silica) or a few tens of W/K m for the best silicon nitrides. For the main part, these electrical insulators are also quite good thermal insulators, and this leads to problems as we shall see below.

12.3.2 Microscopic Order in the Semiconductor

The vast majority of semiconductor components comprise several layers of very high purity monocrystalline semiconductor, of varying thicknesses and with very high crystal quality. In an amorphous semiconductor, the following points should be noted:

- Electron mobility is much lower than in the monocrystal, typically by some 3 orders of magnitude.
- The thermal conductivity is at least an order of magnitude lower.
- The key physical properties represented by the band gap and the behaviour in an electric field are seriously degraded.
- Even when defects are passivated with hydrogen or deuterium, it is impossible to lower the density of deep levels below a threshold that is generally incompatible with the healthy operation of the component.

For all these reasons, which generally lower the performance by 4 orders of magnitude, very few components use amorphous semiconductors at the time of writing.

In a polycrystalline semiconductor, the characteristics of the semiconductor within each grain can be close to those of the monocrystal, but for any component extending over several grains, perturbations arise due to the boundaries between grains. These perturbations remain prohibitive for light-emitting components. In 2008, no commercial light-emitting diode, and hence also no laser diode, was achieved on a polycrystalline semiconductor.

For signal processing functions, by using polycrystals with large grains ($\gg$ 10 μm) and a high level of crystallographic alignment between grains, not to

mention a delicate process of technological optimisation, it is nevertheless possible to obtain a level of performance that is not too far removed from that of the same type of component on a monocrystal, with a degradation by less than two orders of magnitude. In this way, it has been possible to use polycrystalline silicon films for pixel addressing components on recent large area screens and signs. The thermal conductivity of the polycrystal with large oriented grains can approach that of the monocrystal, with a level of degradation that does not exceed a factor of 2 to 3. For example, several suppliers offer polycrystalline SiC manufactured by CVD with a claimed thermal conductivity at 300 K greater than 250 W/K m, which is more than half that of the best SiC monocrystal.

The second example of large scale use of polycrystalline silicon as active material is provided by photovoltaic cells. In this case, the choice between monocrystal and polycrystal results from a compromise between technical, environmental, and economic considerations.

In the monocrystalline semiconductor, doping, and more generally the presence of impurities in the crystal lattice, can lower the thermal conductivity owing to dispersion in the atomic mass, which perturbs the propagation of vibrational modes. With typical dopants, whose atomic mass does not differ too greatly from that of the replaced element, a doping level of about 100 ppm is required for the degradation to reach a few percent. However, cases are known where a contamination by a few ppm by elements with very different atomic masses leads to an unexpected drop of the thermal conductivity by a factor of 2.

In contrast to this, in a very pure crystal, a reduction of the isotopic dispersion can improve the thermal conduction quite significantly. In the literature, one finds conductivities of 200 W/K m @300 K for enriched silicon, i.e., around +30%, and more than 3 000 W/K m @300 K for enriched diamond, which corresponds to about +50% as compared with monocrystals having the natural isotopic distribution.

It now looks possible to use enriched silicon to improve the cooling of certain components. There is at least one new company in the US offering products and services based on this idea.

12.3.3 The Substrate

Wafers

The main object transformed on the production line in the microelectronics industry occurs in the form of wafers, usually round, but sometimes rectangular, e.g., for screens and other visualisation devices. For components designed to control high powers, e.g., in the megawatt range, a component requires a complete wafer and production can be considered to be individual. For example, a PIN rectifier component in the range 5 kV–1 kA occupies a complete wafer of diameter 3 inches (76.2 mm). For all other types of component, production is collective and each processed

and divided wafer provides anywhere between a few tens of chips (large processors) and a few thousand chips (small LEDs).

The increased area of the wafers is due to gain in productivity driven by international competition. At the present time, a large part of world chip production starts from wafers with diameters of 300 or 200 mm. Only a small minority of production units work with wafers of diameter 100, or even 76 or 50 mm.

The full production process includes several hundred stages, each involving several manipulations of the wafer. Current production standards regarding yield rule out a rejection rate above 1%. The wafer must therefore be tough enough to withstand several hundred manipulations. It must also be rigid and elastic enough to allow the required lithographic accuracy, often submicron in 2008, and this imposes a minimum thickness that depends on the material.

As an example, the thickness of silicon wafers of diameter 300 mm used for CMOS chips is around a millimeter, while that of 150 mm wafers for components in the 600 V (1–10 A) range is only about 250 μm. The investment required to set up a new and competitive production unit for VLSI CMOS integrated circuits is well above 10^9 euros. In such a production line, every instrument is optimised for a given diameter and thickness of the substrate. Any change that might be required in the substrate characteristics by a modification of international standards would thus encounter a considerable economic barrier.

For other components, made on wafers with smaller diameter and with less stringent lithographic specifications, the equipment used is generally recycled from former VLSI CMOS production lines. This considerably reduces investment costs and leaves room for variations in the substrate. For example, glass substrates are now used to make networks of passive components.

Active Layer

With the exception of components for use at very high voltages, the active layer of the component, necessarily made from a highly ordered monocrystal with strictly controlled doping, is extremely thin, between a fraction of a micron and a few tens of microns. This monocrystal with its controlled doping can be made from a substrate that is itself monocrystalline. This is the least complex and most widely used solution. However, it has the disadvantage of making it very difficult or even impossible to optimise with regard to heat transfer. The kind of material used for the active layer is imposed by the function it must carry out and the current state of the art.

For example, in 2008, in order to make a light-emitting diode emitting in the blue, the active layer must be a monocrystalline stack of III–N compounds in the AlGaN/InGaN family, produced by epitaxy. The state of the art in epitaxy means that this stack can only be achieved on a monocrystalline GaN, sapphire (hexagonal Al_2O_3), or SiC substrate. For red emission, stacks of AlGaInP compounds are required, and for the infrared, AlGaAs or AlGaSb compounds. At the present time, despite many failed attempts and an international scandal around the year 2000

concerning purportedly miraculous solutions, such stacks can only be achieved on monocrystalline substrates of the same crystal system, matched to the interatomic separation of the compounds in the stack, which considerably limits the number of possible configurations.

For data storage and processing, signals must be transported at high frequencies with minimal energy loss. The ideal would be to use a thick insulating substrate with a low dielectric loss coefficient. We are coming close to this in monolithic UHF integrated circuits which use semiconductor circuits made insulating by the deliberate introduction of centers generating deep levels in the band gap, viz., carbon-doped GaAs endowed with EL2 levels for circuits operating between 2 and more than 60 GHz, iron-doped InP for still higher frequencies, vanadium- or titanium-doped SiC endowed in some cases with specific crystal defects for future generations of power circuit operating between 1 and 20 GHz. The thermal conductivities of GaAs and InP are very poor (0.5 and 0.7 W/K cm).

Diamond

For these applications, from the thermal and dielectric standpoint, the designer's dream would be a monocrystalline diamond substrate with a thermal conductivity of 2000 W/K m. Progress in the industrial development of bulk monocrystalline diamond is slow, probably also hindered by the interests of certain stakeholders in the natural diamond market. Despite the fact that microwave plasma reactors and techniques are available, crystal growth remains slow and difficult. The only seeds of high enough crystal quality are certain natural diamonds, the most perfect, the rarest, and consequently the most costly. Lateral growth and increase in area of the crystals are not yet well controlled. Samples of diameter 8 mm have been mentioned in the press, and rumours have been going around of 25 and even 50 mm wafers.

But even if monocrystalline diamond wafers were available, one would still have to make the monocrystalline active layer of the desired component on this substrate. Epitaxy cannot directly deposit the silicon (or GaAs, InP, SiC, GaN, etc.) monocrystal on diamond, even if it is monocrystalline. Although silicon and diamond have the same crystal system, the interatomic distance is different, viz., 154 pm for diamond and about 236 pm for silicon. Any attempt to anchor one crystal lattice on the other results in a very high density of extended crystal defects, i.e., dislocations and stacking faults. The only known exception at the present time may be the epitaxy of certain III–N compounds (GaN, InGaN, AlGaN), which may be possible on monocrystalline diamond.

There now exist molecular bonding techniques that can be used to assemble very clean and previously 'activated' surfaces of various materials.

While monocrystalline diamond with large area and accessible prices remains largely a dream for future decades, polycrystalline diamond already exists in the form of wafers of diameter up to 125 mm. The price of such remains prohibitive (around 100 kEUR for a wafer of diameter 125 mm) mainly due to the low demand, combined with the near monopoly maintained by a very small number of suppliers

and the thwarting of certain potential uses by a particularly aggressive policy of intellectual ownership by some industrial groups who have succeeded in strangling the sector without having ever really managed to launch a market for applications resulting from their own work.

Many groups around the world are able to grow polycrystalline diamond wafers, but not many groups or consortiums possess the means required for polishing, processing, and cutting into chips, essential operations for establishing this type of substrate as a new raw material in the microelectronics industry. It is still very difficult to polish diamond, especially its $\langle 111 \rangle$ face, which is the hardest, and few groups have this capability. In Europe, the main companies are the world leader Element Six, a subsidiary of De Beers and Umicore, and the new company Diamond Materials, progeny of the Fraunhofer Institute.

Polycristallinity leads to the presence of grain boundaries and different hardness from one grain to another, hence a roughness that constitutes an obstacle to molecular bonding, since the latter is extremely demanding with regard to the flatness of the surfaces to be set together. (An rms roughness well below the nanometer is needed.)

Even assuming perfect molecular bonding, one still has to face the problem of differential thermal expansion. Most components must be able to withstand a chip temperature range from -50 to $+150$°C. The stresses due to differential thermal expansion over a hundred degrees (several ppm/K $\times$ 100 K implies several hundred ppm) are considerable for chips of millimetric dimensions. Several gigapascals can be reached, well above the dislocation threshold of typical semiconductor monocrystals.

Historically, there have been computers (made by CRAY) using GaAs chips mounted on polycrystalline diamond substrates. The problem of differential expansion was rather limited here by the fact that the chips where mounted on a metal intermediate and operated at an almost constant temperature, in liquid nitrogen.

To end this discussion, note that the electrical conduction resulting from high boron doping ($> 1^{E}20$ cm^{-3}) raises the possibility of one day making electrically conducting diamond substrates, with comparable electrical resistivity to highly conducting silicon or monocrystalline SiC substrates (of the order of 10–20 milliohm cm). The efficiency of high boron doping should not be taken to mean that the problems of doping diamond as a semiconductor have been fully resolved. To our knowledge, this problem remains, and there are only a few small scale applications for which diamond can be used today, e.g., detection of nuclear radiation, or detection of hard UV radiation. The lack of an efficient doping process remains an obstacle, still to be overcome in 2008, to the development of electronic components using active layers made from semiconducting diamond.

Several programs are underway in France to validate a new type N doping process proposed by the French state research organisation (CNRS) and to explore the possibilities of phosphor doping. There are also some ideas for making diamond components without the need to dope the material.

What has been said here for diamond applies also to *sp*3 boron nitrides, and the cubic and hexagonal variants of wurtzite, which remain largely unexplored and may eventually provide an interesting alternative to diamond.

Silicon Carbide

While diamond substrates remain unavailable, there is a monocrystalline substrate whose thermal conductivity at room temperature lies close to that of copper. This is silicon carbide, which can be electrically conducting when doped, e.g., N-doped with nitrogen, and can be semi-insulating with a typical resistivity of around $1^{E}12$ ohm cm at 300 K). Available in the form of wafers with diameters up to 100 mm, it is already used in extreme applications where it proves essential to obtain maximal performance, in particular for the development of new UHF power transistors using GaN, but also for the fabrication of LEDs emitting in the spectral range from the green to the near ultraviolet.

It remains very difficult and costly to make (around 3 kEUR per wafer) and is only available from a few suppliers around the world (Cree in the USA, Sicrystal in Germany, an offshoot of Siemens/Osram/Infinenon, Norstel in Sweden, and a few newcomers such as Bridgestone, II–VI, and Caracal). Growth of high quality monocrystal is slow (< 1 mm/hr) and requires temperatures close to 2 000°C. Processes for increasing the size of the crystals are complex and delicate, and they are highly sensitive to temperature gradients. They are used to make certain light-emitting diodes and some components for energy control, in particular, fast Schottky rectifiers which are already commercially available. In 2008 it would be difficult to imagine this type of substrate being used to make components for storing and processing data that might replace the CMOS, even if the epitaxial processes for III–N compounds on SiC could provide a way of making complementary transistors that look likely to outperform a MOS of the same dimensions on silicon.

On the other hand, high purity polycrystalline SiC exists in the form of ingots and wafers of diameter up to more than 300 mm, at a reasonable price, and with a fairly similar thermal conductivity to the monocrystal (250–350 W/K m as compared with 450–500 W/K m). The insulating or conducting nature of this material, still difficult to reproduce in a reliable way, might be better controlled, if applications so required, by controlling the purity and concentration of one or two well chosen impurities. A program entitled HYPHEN, coordinated by Picogiga–Soitec and supported by the European Community, with the participation of Thales and IEMN, has begun to validate the use of this SiC material to carry a composite substrate. In this project, the validation of the composite substrate is achieved by demonstrating and characterising UHF power transistors.

Composite Substrates

Being able to deposit silicon MOS transistors on an insulating sublayer, to remove the need for P/N junctions to insulate the individual components, represents a significant step forward, which has justified the development of processes for assembling semiconducting layers on thin insulating films in so-called composite substrates. The leader here is SOITEC. For those component manufacturers who adopt this type of substrate, the gain in performance would correspond to roughly one generation on the Road Map produced by the international association of semiconductor manufacturers (SIA).

In 2008, this insulating layer is still made from silica (SiO_2), which is also highly insulating from a thermal standpoint. Placed right next to the transistor, its effect on the efficient cooling of the transistor via the substrate is all the more radical. A priority for upstream research is therefore to find an insulator with better thermal conductivity. Silicon nitride (Si_2N_4) can exhibit a thermal conductivity up to 50 times better. The replacement of silicon by silicon nitride must be difficult, otherwise it would already have been achieved, bearing in mind that the idea has been in the air for 10 years now. Thin diamond films are also being investigated, and one can imagine other alternatives based on boron or aluminium nitrides.

For the base substrate, the insulating or clearly conducting materials with the highest thermal conductivities already mentioned (diamond, SiC) will be preferred, even if the technological difficulties mean that almost all composite substrates are at present realised on silicon substrates optimised for the cost/quality compromise.

Exploratory research had already been carried out at the end of the 1990s to extend the range of composite substrates by transferring thin monocrystalline films of III–V GaAs or InP semiconductors onto an insulated substrate. The structural quality of the resulting monocrystal was insufficient to obtain UHF components with high enough performance, and the reliability of optoelectronic components made on this type of substrate was much too low to allow industrialisation. Further research is now underway to demonstrate the reliability of substrates using a thin monocrystalline film of GaN with a view to making large light-emitting components on less costly substrates than SiC or sapphire. Many delicate problems remain to be solved and the success of this program is not yet guaranteed.

Using the same family of assembly techniques as for composite substrates, there are now electronic circuits on glass and processes for transferring whole chips onto various types of substrate. Some applications leading to large series give rise to specific development of substrates that are perfectly adapted to the context of the application.

12.4 Problems and Predictions for the Main Chip Types

12.4.1 Components for Controlling Electrical Energy

Low Frequencies

At low frequencies, these components are used in two stable states which we may call the on state and the off state. The transition is made from one state to the other by a switching phase. In the stable states, energy dissipation is lower, because one of the two variables, current and voltage, remains small.

In the switching states, which last for only a fraction of the operating cycle, the current and voltage are high, and the power dissipated is maximal. In an optimised system, most of the energy losses occur during switching. On the one hand, the energy dissipated at each switching will be lower if the switching is faster. However, fast switching raises electromagnetic problems, in particular, by causing interfering radiation over a broad spectrum, and voltage surges ($-L\mathrm{d}I/\mathrm{d}t$) within the inevitable inductive terms due to connection tracks and cables.

In the on state, dissipation in the chip arises mainly from the series resistance in the semiconductor and possibly also from the metal/semiconductor contacts. On the one hand, any rise in temperature will cause the electron mobility, and hence also the electrical conductivity, to drop, which will increase the dissipated power still further. But in addition, any rise in temperature will also cause the thermal conductivity of the semiconductor to drop. The system is unstable as a result of this double positive feedback. Efficient evacuation of calories and reduction of the temperature of the active layer, even at the cost of extra external energy consumption/dissipation, are essential features when designing the architecture and environment of the component.

In the off state, the higher the required breakdown voltage, the purer the material must be, and hence the lower the doping level. For any semiconductor, there is a temperature threshold for which the intrinsic carrier density meets the extrinsic density due to doping. For a material with low doping and narrow band gap, this threshold can be low. For a material like silicon and a breakdown voltage of 1 kV, corresponding to doping levels below 10 ppb, the maximum admissible temperature at any point of the chip in the off state is around 150°C.

Beyond this threshold, for any local temperature rise, mobile charges resulting from thermal generation of electron–hole pairs can reduce the local resistivity of the material, increasing the current circulating in the hot spot, and thereby further increasing the heating in a thermal runaway effect.

To be efficient, cooling must be implemented uniformly across the whole cross-section of the current flow. As far as possible, the aim here would be to arrange locally for a negative feedback that could focus cooling capacity across a whole region in which a temperature rise had been detected.

For a given breakdown voltage, the thermal sensitivity of components for power electronics will be reduced by introducing semiconductors with a broad band gap,

such as silicon carbide and/or nitrides of gallium–aluminium–indium, then diamond and/or boron nitrides which can withstand much higher electric fields than silicon. Apart from the gain in electrical performance, the thermal stability will be greatly improved, along with the reliability, a gain strengthened by the high thermal conductivity of these materials.

High Frequencies

As mentioned in the introduction, at high frequencies, typically above 10 MHz, there are no switches fast enough to synthesise signals by switching and filtering. It is the transistors that modulate a current according to the desired shape of the signal. Voltage and current are present simultaneously in the semiconductor for a significant fraction of the working cycle. Efficiencies are lower than for the switching states.

Some applications run at a single frequency and with constant output load. This is called narrow band protected output operation. In this case, the reactive elements of the circuits, capacitances and inductances, can be tuned by conjugate elements, the load impedance can be matched to the optimal output impedance of the transistor, and differential tuning can be achieved over several output harmonics of the transistor (class E and F operation). Applying all these devices, the simultaneous presence of voltage and current in the transistor can be avoided. The energy efficiency can then exceed 60% up to 10–15 GHz and even 80% below 3 GHz.

The same kinds of ploy cannot be efficiently implemented for broad band operation of a UHF semiconducting power component. The efficiency falls off rapidly with the frequency and the pass band, and also with linearity constraints. Furthermore, performance also drops off rapidly with temperature, once again due to degradation of electron mobility. Reliability is also degraded by temperature. The main degradation mechanisms involve migration of metals in the semiconductors, and increased leakage currents at the potential barriers along the path of the electrons in the current-blocked state.

Chip cooling is complicated by the electromagnetic constraints laid down by impedance matching. Metal or insulating objects cannot be inserted, removed, or displaced in an arbitrary manner in the vicinity of the chip. Each geometric modification requires electromagnetic reoptimisation and redesign.

Two cases are particularly complex. The first involves base stations for mobile phones, where transistors deliver around a hundred watts at 1–2 GHz. The base station serves around 1 000 subscribers, and hence hundreds of communication channels in parallel. Even when the spectrum is shared between several amplifiers, each amplifier must process dozens of channels without interference, ideally with perfect linearity. With recent modulation codes, in order to pass a lot of data through a narrow channel, the signals are modulated in both amplitude and phase. The linearity requirement then becomes extreme. Linearity correction systems are already in place by predistortion of transistor control and post-correction. Despite this sophisticated arrangement, the only way to approximate the required specifications is to run the transistors with small signals well below their maximum capabilities, in a

regime where the energy conversion efficiency is very low, often less than 20%. To emit 100 W, at least 400 W must be dissipated.

And here begins the ultimate complication. Any fluctuation in the dissipated power will cause the temperature to vary, and with it the physical properties of the semiconductor. It follows that thermal fluctuations lead to a non-steady state response from the transistor, and this results in effects on the linearity with characteristic time constants that range over many orders of magnitude, and which must be taken into account for linearity corrections. But to make matters worse, these base stations are located at the top of a mast, exposed to the sun, and the outside temperature can sometimes reach 60°C, while the inside temperature at the base of the chip must not generally exceed 70°C. This leaves an available margin of only 10 K for evacuating hundreds of watts of heat and maintaining an internal temperature stable to within a fraction of a degree under extremely fluctuating conditions of electrical load. Using air cooling, a large surface area would have to be attached to the mast with all the resulting problems of wind resistance, aesthetics, and the radiation pattern of the antennas. Using water cooling, the liquid must be raised to the top of the mast and then brought down again, avoiding leakage and bubbles, a whole problem in itself. This explains the complexity and cost of base stations, for which the world turnover is close to that of portable terminals, even though the latter outnumber them by a thousand to one.

For these applications, research is always on the lookout for improvements in the semiconductors themselves, or any type of switching that can switch tens of electrical nanojoules in several tens of picoseconds with low losses, not to mention techniques for cooling chips, components, cards, and modules.

The other extreme application concerns broad band amplifiers. These amplifiers are used in military electronic war applications, but also, and more and more commonly, when the same amplifier is required to process the numerous radio signals of everyday life in the developed countries, viz., GSM, UMTS, Blue Tooth, WiFi, Wimax, their successors, and their future developments.

Broad band implies reductions in the reactive elements, viz., capacitances, inductances, transmission lines, and contacts, all of which leads to compact architectures. It also implies impedance matching and imperfect tuning, and intrinsic energy efficiencies limited by the characteristics of the semiconducting material. With compactness and high dissipation, the dissipated heat density really takes off. With the advent of the III–N family of semiconductors, absolutely unheard-of values are reached, e.g., tens of milliwatts per square micron locally on the chip, and tens of watts per square millimeter at the base of the chip. Such densities cannot be dealt with using conventional techniques which depend solely on heat conduction by the substrate. A revolution is needed at all levels in the system architecture. Regarding industrial competitivity, only those who implement this revolution in good time are likely to survive. Those who can maintain amplifier performance with optimised thermal structure will also dominate the markets for systems using such amplifiers.

12.4.2 Processor and Memory

At the present time, data storage and processing functions are carried out by highly complex CMOS circuits which evolve in accordance with Moore's law, now internationally accepted as a guide for the development of future generations of components, at least until 2020, and which suggests that processor complexity will double on average every two years.

The heat equation for this type of chip follows from several well-defined phenomena. On the one hand, the elementary operation carried out by a MOS transistor is a charging and discharging cycle with an electrical capacitance C comprising gates of other transistors and data transmission lines. If V is the supply voltage of the chip, the energy stored in this capacitance is close to $CV^2/2$. In the present working mode of memories and processors, this energy is lost in each cycle and transformed into heat in the channel of the MOS transistor. To minimise such energy losses and heat dissipation, several ideas are implemented in parallel.

Reducing the Capacitance

Size Reduction

Reduction of the lateral dimensions (minimal gate width around 45–60 nm in industrial products in 2008) reduces the capacitance C, a beneficial effect. It also makes the transistor faster. Reducing the thickness of the gate oxide (around 1–2 nm in 2008, depending on the manufacturer), which is essential for maintaining the functionality of small transistors, raises enormous difficulties. Some manufacturers around the world have apparently solved these problems for the years 2008–2015, at the cost of completely changing the insulating material (to use high permittivity materials) and the production process. Beyond 2015, uncertainties remain. In any case, reducing the thickness of this insulator and increasing its permittivity work to some extent against the gain achieved by reducing the lateral dimensions.

Silicon on Insulator

Silicon on insulator provides a very interesting way of reducing unwanted capacitances due to the insulating junctions of the source and drain electrodes in the transistors. There are commercially available substrates comprising a thin film of monocrystalline silicon assembled on a silicon substrate with an intermediate layer of silicon oxide. These substrates are already used by some processor manufacturers who claim to be able to gain roughly one generation, i.e., 2 years of technological development, with respect to calculation performance. Still more would be gained by using a true insulator with low dielectric losses as base substrate, but the substitution or layering of semiconducting materials is problematic.

Reducing the Supply Voltage

To strengthen, and limit the peak value of the electric field within the structure, it is also desirable to reduce the supply voltage. This is indeed what has been happening. It has dropped from around 5 V in the 1970s to 1–2 V by 2008, depending on the type of circuit. Three difficulties are encountered when attempting to reduce it further:

- The energy stored in the elementary cell must be big enough to ensure that it will withstand noise, in particular noise due to cosmic rays, which is beginning to be identified as a cause for concern, even at low altitudes, but still more so at higher altitudes.
- The value of the voltage must be compatible with the transconductance available in the transistors. The latter no long progresses in direct proportion to size reduction.
- The voltage must remain well above statistical technological scatter in the threshold voltages of the transistors. Despite a huge effort on all levels, the greater number of elementary cells progresses faster than the tightening of standard deviations achieved by improved manufacturing processes.

As a consequence, the supply voltage of successive generations of chips has been falling very slowly, and will continue to fall very slowly. We have reached the point where we must choose between a simpler architecture with higher voltage and a lower voltage accompanied by a more sophisticated architecture including several control levels with error detection redundancy to overcome problems of noise, electromagnetic interference, and technological scatter. In both cases, a limit is about to arise in the reduction of the energy handled by an elementary function.

Energy Recovery

In today's circuits, each time a capacitive cell is discharged, the energy used to store the data temporarily is lost, dissipated in the form of heat in a transistor channel designed to allow discharge. But it may well be that this environmentally unfriendly practice is not inevitable. It should not be impossible to recover this energy in a similar capacitance, rather than simply dissipating it. In principle, it would suffice to connect the two capacitances by an inductive bridge for a time equal to the half-period of the resulting oscillating circuit. Such an operation leads to a simple energy transfer from the initially charged capacitance to the initially uncharged capacitance.

Another idea also uses an inductive connection switched on for a half-period, but this time placed temporarily in parallel with the capacitance whose state is to be changed. When carried out in an optimal way, a cycle of this kind changes the sign of the voltage across the capacitor terminals, while conserving the stored energy. By making a small change, viz., fixing the reference voltage at half the supply voltage, the resulting operation is indeed an inversion of the stored binary digital content, without massive and systematic loss of energy. By doing this, in order to save energy,

we increase the complexity of the elementary cycle, and hence also the complexity involved in managing sequences.

Beyond a certain level of complexity, one must choose carefully between many technological options. Those enhancing performance without due consideration for heat transfer and energy consumption will no longer be viable. The choice will be between several strategies for economising consumed and dissipated energy. It will be important to envisage, simulate, and choose between these options by 2020.

12.4.3 Light-Emitting Components

The energy efficiency of light-emitting diodes is increasing rapidly. The switch to direct band gap semiconductors in the years 1985–2000 was a decisive step in this process, bringing an improvement by at least an order of magnitude in the energy efficiency. Progress is still being made, by optimised engineering of heterostructures and better organised recovery of the light. In 2008, the efficiency of a blue GaN diode is some 60% at research level and 15% in commercial products. There is still considerable heat dissipation in this component, and its efficiency drops with the temperature of the active layer, but the power density to be dissipated is not as high as for data processing components, and conventional cooling solutions by conduction through the substrate should prove adequate here.

One specific difficulty lies in the semiconducting materials themselves. These are often ternary or quaternary III–V alloys. In such alloys, the elements of at least one of the families, i.e., III elements or V elements, position themselves at random on the sites of the corresponding sublattice. For substituted elements with very different masses, disorder can lead to a spectacular reduction of the thermal conductivity. For example, it has been known since the 1990s that the thermal conductivity of the disordered compound $Ga_{0.5}In_{0.5}P$ epitaxied on GaAs, used for red light emission and UHF emission in portable telephones, is at least three times lower than that of the constituent binary compounds InP and GaP. Unfortunately, regarding the structures of most successful LEDs, there are few or no alternatives to the optimised stackings currently used. This means that the component must often be taken as it comes with low thermal conductivity.

Concerning laser diodes, which are specially arranged to obtain self-stimulated emission, there is a problem of control and uniformity of temperature, especially for narrow band diodes such as distributed feedback (DFB) laser diodes. Any temperature fluctuation, even local, shifts the optical gain curve of the semiconductor stacking. For a simple laser diode with no external filter, subjected to a uniform temperature rise, a shift and slight broadening of the emitted line are observed. For a non-uniform temperature distribution, the threshold current of the laser diode increases and its spectral purity is degraded. In some cases, jumps can occur in the guiding mode within the semiconductor cavity, sometimes accompanied by a splitting of the emitted line. For a laser diode with an external filter (DFB), the gain curve of the semiconductor can sometimes shift far enough from that of the filter to

cause a spectacular shift in the output power, or even a disconnection of the laser oscillator.

These phenomena can themselves alter the heat dissipation distribution in the diode and sometimes lead to destructive blockages. It is essential when designing optronic systems to optimise the cooling of laser diodes. Such optimisation naturally involves reducing the average temperature, but also the temperature gradients, and careful surveillance of changes in the characteristics of the laser which may be influenced by the thermal configuration of the chip, especially in cases where the optical load at the laser output is subject to variations and where a significant part of the power emitted by the laser might be sent back into the cavity.

12.4.4 Trends in Heat Transfer Features of Semiconductor Components in the Coming Decades

To begin with, thermal conduction through the substrate will be improved by changing the material in the substrate and by thinning down the regions with lowest thermal conductivity. Concerning bulk materials, as already mentioned, the future lies in materials with high thermal conductivity, such as diamond, SiC, and boron and aluminium nitrides.

To solve the problem of differential expansion, there will be no choice but to insert intermediates between bulk objects with different properties, e.g., diamond and silicon, GaAs, InP, GaN, etc. To make the match, bundles and lattices of thermally conducting micro- and nanofibres, sheets, and tubes, electrically insulating or conducting depending on the application, would appear to present the best approach, even though significant technological difficulties can be foreseen in carrying out such an insertion collectively, at an acceptable process temperature, without compromising the electrical performance or overburdening production costs.

It will soon become impossible to cool the chip solely via the substrate. Heat must also be evacuated through the free surface, the one currently used for electrical connections. The electrical insulators now used on this face of the chip are also silica based thermal insulators. They will have to be replaced by thermal conductors, and it may also be necessary to organise connection networks with mixed electrical and thermal functionalities, perhaps using two-phase microloops or micro- or nanocaloducts.

In the long term, wired electrical connections and supplies will very probably be replaced by remote optical systems, with all the corresponding simplifications in the general infrastructure of the chip, but with complications to local functional units.

To transfer calories from the chip to less dense regions of the system, two-phase (evaporation/condensation) loops would appear to hold the best answer. Passive loops will quickly reach their limits in terms of power density, and there will be no choice but to move to active loops incorporating pumps and compressors, probably throughout the architecture.

The development of components for active loops with the required performance, compactness, and reliability certainly represents a major issue for system competitivity from microelectronic to macroscopic features of the architecture. One way to meet this challenge will be to develop electrical/magnetic/mechanical micro- and nanosystems.

References

1. J.P. Holman: *Heat Transfer*, Metric Editions, Mechanical Engineering Series, McGraw Hill
2. J. Taine, J.P. Petit: *Transferts thermiques. Mécanique des fluides anisothermes*, Dunod
3. F.P. Incropera, D.P. Dewitt: *Introduction to Heat Transfer*, 3rd edn., Wiley
4. V.L. Hein: Convection and conduction cooling of substrates containing multiple heat sources, The Bell System Technical Journal, Vol. XLVI, Oct 1967, no. 8, pp. 1659–1678
5. F. Dhondt: Modélisation electrothermique des transistors bipolaires à hétérojonction (TBH) pour les applications de puissance à haut rendement en bande X, PhD. Thesis at the University of Lille, France
6. R. Mehandru, S. Kim, J. Kim, F. Ren, J.R. Lothian, S.J. Pearton, S.S. Park, Y.J. Park: Thermal simulations of high power, bulk GaN rectifiers, Solid State Electronics **47**, 1037–1043 (2003)
7. S. Orain: Etude théorique et expérimentale des phénomènes de conduction thermique dans les matériaux diélectriques déposés en couche minces. Application aux dépôts d'oxyde, PhD Thesis at the University of Nantes
8. M. Gerl, J.P. Issi: *Traité des matériaux*, Vol. 8, *Physique des Matériaux*, Presses Polytechniques et Universitaires Romandes, p. 702
9. J.M. Ziman: *Electron and Phonons. The Theory of Transport Phenomena in Solids*, Oxford University Press (1960) p. 545
10. R. Breman: *Thermal Conduction in Solids*, Oxford University Press (1976) pp. 66–69
11. J.E. Parrot, A.D. Stukes: *Thermal Conductivity of Solids*, Pion, London (1975) pp. 44–122
12. J.C. Lambropoulos, S.D. Jacobs, et al.: Thermal conductivity of thin films: Measurement and microstructure effects, HTD Thin Film Heat Transfer: Properties and Proceedings ASME **184**, 21–32 (1991)
13. D.I. Florescu, V.M. Asnin, L.G. Mourokh, F.H Pollack, R.J. Molnar: Doping dependence of the thermal conductivity of hybrid vapour phase epitaxy grown *n*-GaN/sapphire (0001) using a scanning thermal microscope, *Symposium on Gallium Nitride and Related Alloys*, at the 1999 Fall Meeting of the Materials Research Society held in Boston
14. D.I. Florescu, L.G. Mourokh, F.H Pollack, D.C. Look, G. Cantwell, X. Li: High spatial resolution thermal conductivity of bulk ZnO (0001), J. Appl. Phys. **91** (2) (2002)
15. M.E. Brinson, W. Dunstan: Thermal conductivity and thermoelectric power of heavily doped *n*-type silicon, J. Phys. C Solid St. Phys. **3** (1970)
16. D. Kotchetkov, A.A. Balandin: Modelling of the thermal conductivity of polycrystalline GaN film
17. D. Kotchetkov, J. Zou, A.A. Balandin: Theoretical investigation of thermal conductivity in wurtzite GaN, Mat. Res. Soc. Symp. Proc. **731**, W5.11 (2002)
18. B.C. Daly, H.J. Maris, A.V. Nurmikko, M. Kuball, J. Han: Optical pump-and-probe measurement of the thermal conductivity of nitride thin films, J. Appl. Phys. **92** (7), 3821–3824 (2002)
19. D.I. Florescu, V.M. Asnin, F.H Pollack, A.M. Jones, J. Ramer, M. Schurman, I. Ferguson: Thermal conductivity of fully and partially coalesced lateral epitaxial overgrown GaN/sapphire (0001) by scanning thermal microscopy, Appl. Phys. Lett. **77** (10), 1464–1466 (2000)
20. J. Zou, D. Kotchetov, A.A. Balandin, D.I. Florescu, F.H. Pollack: Thermal conductivity of GaN films: Effects of impurities and dislocations, J. Appl. Phys. **92** (5), 2534–2539 (2002)

21. J. Zou, D. Kotchetov, A.A. Balandin, D.I. Florescu, F.H. Pollack: Effect of dislocation on thermal conductivity of GaN layers, Appl. Phys. Lett. **79** (26), 4316–4318 (2001)
22. Y. Gu, D. Zhu, L. Han, X. Ruan: Imaging of thermal conductivity with lateral resolution of sub-micrometer using scanning thermal microscopy, *Fourteenth Symposium on Thermophysical Properties*, 25–30 June 2000, Boulder, Colorado, USA
23. A.L. Palisoc, Y.J. Min, C.C. Lee: Thermal properties of five-layer infinite plate structure with embedded heat source, J. Appl. Phys. **65** (11), 4438–4444 (1989)
24. A.L. Palisoc, C.C. Lee: Exact thermal representation of multilayer rectangular structures by infinite plate structures using the method of images, J. Appl. Phys. **64** (12), 6851–6857 (1988)
25. C.C. Lee, A.L. Palisoc: Real-time thermal design of integrated circuit devices, IEEE Transactions on Components, Hybrids, and Manufacturing Technology **11** (4), 485–492 (1988)
26. C.C. Lee, A.L. Palisoc, Y.J. Min: Thermal analysis of integrated circuit devices and packages, IEEE Transactions on Components, Hybrids, and Manufacturing Technology **12** (4), 701–709 (1992)
27. K. Kurabayashi, K.E. Goodson: Precision measurement and mapping of die-attach thermal resistance, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A, **21** (3), 506–514 (1998)
28. T.-Y. Chiang, K. Banerjee, K.C. Saraswat: Analytical thermal model for multilevel VLSI interconnects incorporating via effect, IEEE Electron Device Letters **23** (1), 31–33 (2002)
29. J. Park, M.S. Shin, C.C. Lee: Thermal modelling and measurement of GaN-based HFET devices, IEEE Electron Device Letters **24** (7), 424–426 (2003)
30. G.K. Wachutka: Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modelling, IEEE Transactions on Computer-Aided Design **9** (11), 1141–1149 (1990)
31. D.H. Chien, C.Y. Wang, C.C. Lee: Temperature solution of five-layer structure with a circular embedded source and its applications, IEEE Transactions on Components, Hybrids, and Manufacturing Technology **15** (5), 707–714 (1992)
32. H. Iwasaki, S. Yokoyama, T. Tsukui, M. Koyano, H. Hori, S. Sano: Evaluation of the figure of merit of thermoelectric modules by Harman method, Jpn. J. Appl. Phys. **42**, 3707–3708 (2003)
33. D.H. Smith, A. Fraser, J. O'Neil: Measurement and prediction of operating temperature for GaAs ICs, Semi-Therm 86 Symposium, Scottsdale, Arizona, 9–11 Dec 1986, pp. 1–20
34. A.G. Kokkas: Thermal analysis of multiple-layer structures, IEEE Transactions on Electron Devices, **21** (11), 674–681 (1974)
35. D.K. Sharma, K.V. Ramanathan: Modelling thermal effects on MOS *I–V* characteristics, IEEE Electron Device Letters **4** (10), 362–364 (1983)
36. T. Aigo, H. Yashiro, M. Goto, A. Jono, A. Tachikawa, A. Moritani: Thermal resistance and electronic characteristics for high electron mobility transistors grown on Si and GaAs substrates by metal–organic chemical vapour deposition, Jpn. J. Appl. Phys. **32**, 5508–5513 (1993)
37. G.-B. Gao, M.-Z. Wang, X. Gui, H. Morkoc: Thermal design studies of high-power heterojunction transistors, IEEE Transactions on Electron Devices **36** (5), 854–862 (1989)
38. G.N. Logvinov, Y.G. Gurevich, I.M. Lashkevich: Surface heat capacity and surface heat impedance: An application to the theory of thermal waves, Jpn. J. Appl. Phys. **42**, 4448–4452 (2003)
39. V. Szekely: A new evaluation method of thermal transient measurement results, Microelectronics Journal **28**, 277–292 (1997)
40. J.K. Lump: Hybrid assemblies. In: *The Electronic Packaging Handbook*, CRC Press, (2000) 7-1/7-25

Chapter 13
Photothermal Techniques

Gilles Tessier

The first demonstrations of photoacoustic and photothermal effects were made in the nineteenth century, but the development of photothermal techniques did not really take off until the 1970s, in particular through the work of Rozencwaig [1, 2]. Today a broad range of methods can be subsumed under this heading, with the common feature that they use light to produce a thermal excitation. By extension, other methods are included, in which a light wave is used to probe a thermal phenomenon. This chapter will be concerned with the latter, and in particular their application to microelectronic technology. Indeed, insofar as they are non-contact, and generally non-invasive, optical measurement techniques are well suited to many micro- and nanoscale heat transfer problems.

This area has long been dominated by techniques involving measurement of infrared thermal emission in the far field. With the exception of recently developed methods using the near infrared, these methods have reached their limits today because their spatial resolution is not good enough to be applicable to micro- and nanoscale heat transfer. Here we shall review the main optical techniques that have emerged recently to get around these limitations. Many of them use modulation, exploiting the excellent signal-to-noise ratios that can be obtained by lock-in methods, but also spatial confinement of the modulated part of the temperature obtained under alternating conditions.

13.1 Introduction

Given the frequencies and integration densities achieved today, the control of heat transfer phenomena has become a crucial feature in the design of integrated circuits. For example, the packaging of these circuits (casing) is thermally optimised, processors are broken up into several centers to facilitate heat evacuation, elementary transistors and junctions use shapes, concepts, and materials designed to reduce consumption and dissipation to a minimum, and the substrates themselves sometimes

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 389
DOI 10.1007/978-3-642-04258-4_13, © Springer-Verlag Berlin Heidelberg 2009

include vertical structuring, e.g., silicon on insulator (SOI), in which the components are chosen to remove calories as efficiently as possible. Everything from the scale of the printed electronic card (a few cm) down to the scale of the elementary transistor (a few nm) has been calculated to limit power consumption and assist in the dissipation of heat.

Beyond these industrial concerns, many devices currently under research and development also use techniques from microelectronics. MEMS, microfluidic devices, and micro- and nanocalorimeters [3, 4] clearly raise the same kind of problems and require accurate knowledge of local temperatures and heat transfer properties.

13.1.1 Problems Specific to Structures Made by a Top–Down Approach

As can be seen in Fig. 13.1, modern integrated circuits are 3D devices including several levels of metal tracks. Various thicknesses of dielectrics cover active regions, which are usually close to the substrate. In some cases, heat transfer considerations have even led to the adoption of so-called flip-chip configurations, in which the active side of the integrated circuit is solidly bound to a dissipator, so that only the rear face of the substrate is visible. Most heat is thus produced in deeply buried regions that are inaccessible to standard heat measurement methods. At the surface, the only information accessible results from the diffusion of heat through the upper

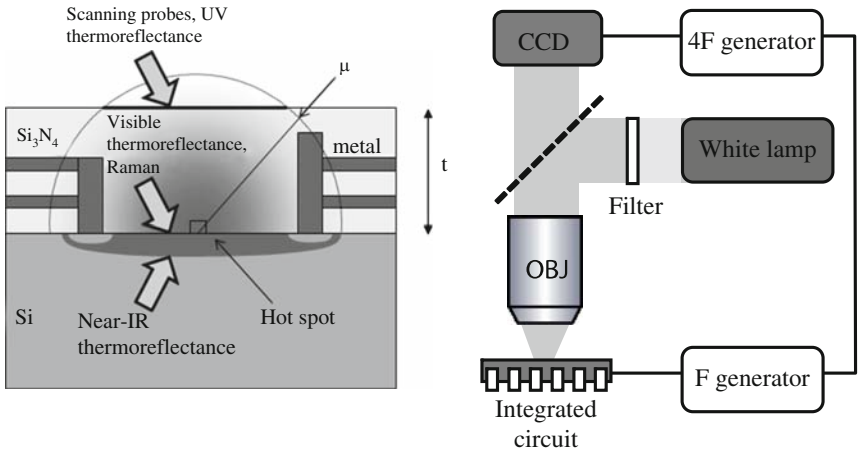

Fig. 13.1 *Left*: Schematic cross-section of an integrated circuit, showing metal layers, interconnects, substrate, and doped regions, together with a passivation layer. Under modulated conditions, the alternating component of the temperature diffuses over a distance μ. *Arrows* indicate regions accessible to the different techniques. *Right*: Schematic view of a CCD thermoreflectance setup

layers. The maximal resolution is therefore at best of the order of the thickness of the intervening layer, i.e., from several hundred nanometers to several micrometers.

Non-contact optical techniques like Raman spectroscopy [5], photoluminescence [6, 7], and thermoreflectance, can provide a way of determining the temperature of these buried items, provided that the materials are transparent to the wavelength used.

13.1.2 Thermoreflectance and CCD Cameras

Thermoreflectance is a non-contact optical method exploiting local reflectivity variations induced by heating to deduce a temperature measurement, and it can be used at practically any wavelength. Diffraction-limited spatial resolutions of the order of $\lambda/2$, i.e., 200–500 nm, can thus be obtained. This measurement can be made by scanning a focused laser beam across the sample surface [8,9], or by using a detector array to obtain a simultaneous measurement at several points. The immediate advantage of the CCD is obviously to reduce the image acquisition time by multiplexing the measurement. The price to pay is generally a reduction in the signal-to-noise ratio owing to the limited number of photons that can be detected (capacity of the CCD wells), and the resulting increase in the relative contribution of photon noise. However, even more important than speed, the main advantage of CCD techniques is probably the possibility of using spatially incoherent sources, since it is no longer necessary to focus the beam in order to obtain good resolution. As a result, almost the whole light spectrum becomes accessible using filaments, arcs, or light-emitting diodes.

As shown in Fig. 13.1, this well understood technique can be used in the visible to measure temperatures through several micrometers of transparent layers with submicrometer spatial resolution and sensitivity generally less than one kelvin. In the ultraviolet, the encapsulating dielectric layers are opaque and thermoreflectance can measure the surface temperature. Finally, in the near infrared, thermoreflectance provides a useful solution for carrying out temperature measurements through silicon substrates, which are transparent in this region of the spectrum. After describing the principles of thermoreflectance measurements using a CCD camera, the rest of the chapter will be devoted to the main wavelength ranges currently under investigation, i.e., the visible, the ultraviolet, and the near infrared.

13.2 Thermoreflectance Imaging

13.2.1 The Underlying Phenomenon

When it reflects on a surface, the amplitude and phase of an incident optical wave are modified. These modifications, depending as they do on the surface, are full of information, which is exploited in ellipsometry, for example. The presence of a variable thermal phenomenon will lead to a variation in the reflection coefficient

$$\mathscr{R}(T) = \left[R_0 + \Delta R(T)\right] \mathrm{e}^{\mathrm{i}[\varphi_0 + \Delta\varphi(T)]} .$$

Changes in the optical phase, which can be measured using interferometric techniques, are related to motions of the surface (especially those induced by expansion), but also by modifications in its optical response. On the other hand, amplitude variations are only related to changes in the permittivity ε, or the index n, which are functions of the temperature. Phenomenologically, this dependence is often treated as linear, whence one has simply

$$\Delta R = \frac{\partial R}{\partial T} \Delta T .$$

This dependence is common to all materials, but covers very different physical processes. Unfortunately, apart from the materials commonly used in optics [10] and certain semiconductors, it is difficult to find reliable values in the literature. For example, in some transparent dielectrics, a large part of this dependence can be explained using a simple electrostriction model to calculate $\varepsilon(T)$, $n(T)$, or $R(T)$, taking expansion into account [12].

In metals, several phenomena modify the optical response as a function of temperature [13, 14]:

- Expansion reduces the plasma frequency and shifts the energy levels, leading to a shift in the Fermi level. This may be affected by constraints imposed by the substrate. The Fermi level may also increase with temperature, but its effect should be small for moderate temperature changes.
- The gap in the Fermi distribution broadens, and this affects certain interband transitions.
- The increased phonon population reduces electron relaxation times, and shifts the energy bands through electron–phonon interactions.

In metals adequately described by a Drude model, such as aluminium, a reasonable value for the coefficient $\partial R/\partial T$ can be obtained by ab initio methods.

Finally, in semiconductors, it is mainly the temperature dependence of the gap that leads to variations in $R(T)$, and hence in $\partial R/\partial T$.

Furthermore, just as ε, n, and hence R depend strongly on the illumination wavelength, the coefficient $\partial R/\partial T$ varies significantly with λ. In many cases, an optimal measurement sensitivity can only be obtained in thermoreflectance by carrying out a full spectroscopic study.

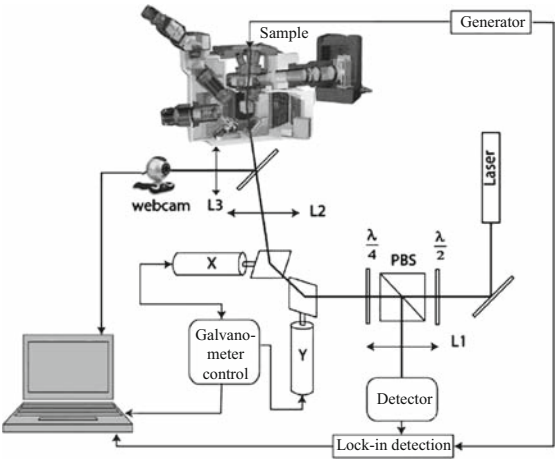

Fig. 13.2 System using a galvanometer mirror to displace a laser beam focused at the surface of the tested circuit. Taken from [8]

13.2.2 Measurement Methods

Generally, this coefficient is of the order of 10^{-3}–10^{-5} K^{-1} for most solids. Owing to the noise intrinsic to most light sources, such small changes in R cannot be measured under direct current conditions. Supplying the circuit at a frequency F causes heating at the same frequency if the signal has a nonzero average. Using a single detector, the measurement can then be carried out with a lock-in amplifier detection at frequency F. This technique can only obtain images by a generally rather long scanning process. Recently, galvanometer mirrors have been used to improve the acquisition rate quite significantly, while benefiting from the excellent dynamic range of the photodiodes [8]. The setup shown in Fig. 13.2 can achieve sensitivities of the order of 10^{-5} in $\Delta R/R$, for a measurement time of the order of 0.2 ms per point. This performance makes galvanometric scanning a good compromise between the sensitivity reached by single beam techniques ($\Delta R/R \approx 10^{-6}$ for 1 s/pixel) and the high acquisition rate of CCD techniques ($\Delta R/R \approx 10^{-4}$ for 0.1 μs/pixel).

With detector arrays (cameras), one can use as many lock-in detections as pixels [15]. In practice, this solution has been implemented for an array of $16 \times 16 = 256$ pixels, each equipped with its own analogue-to-digital conversion electronics. Thanks to rapid data transfer, these data undergo fast Fourier transform (FFT) to extract the component modulated at the excitation frequency. This solution, although limited to a few hundred detectors due to the volume of data to be transferred, is very efficient in terms of signal-to-noise ratio. To increase image definition, an aperture smaller than the pixel can be displaced in front of each detector. To improve the acquisition rate, A. Shakouri and coworkers have used a 16×16 array of holes which, displaced in front of the detector array, serves to refine the image. Using

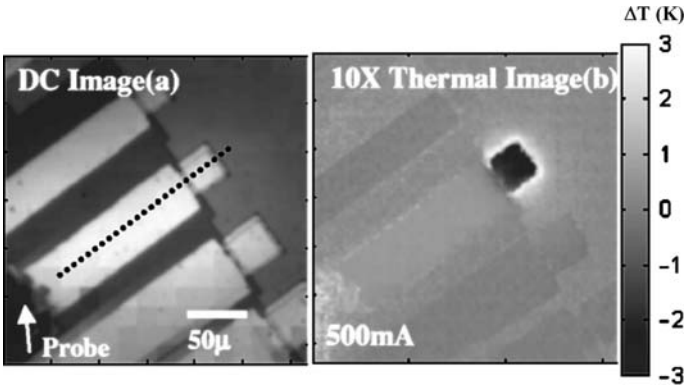

Fig. 13.3 (**a**) Optical and (**b**) thermal images of a micro-Peltier of side 50 μm subjected to a current of 500 mA, obtained using real-time Fourier analysis on the signals from a 16×16 detector array. From [15, 16]

holes corresponding to 1/100 of the pixel area, images of 160×160 pixels have been obtained (see Fig. 13.3).

An alternative is to use the so-called four-bucket method. The camera is then run at a frame rate of $4F$ to make 4 images I_1 to I_4, one in each period of the modulated heating. Assuming the implicit x and y dependence of the relevant quantities (since one is making a 2D image), the amplitude and phase of this modulation are then obtained using the complex $A = I_1 - I_3 + \mathrm{j}(I_2 - I_4)$ [17, 18]. This leads to

$$\left|\frac{\Delta R}{R_0}\right| = \frac{\pi}{\sqrt{2}}\frac{|A|}{I_0} ,$$

if R is modulated sinusoidally, and

$$\left|\frac{\Delta R}{R_0}\right| = \sqrt{2}\frac{|A|}{I_0} ,$$

if R has a square modulation, with $I_0 = (I_1 + I_2 + I_3 + I_4)$ the average of the four images.

On a given material, calculation of the image $\Delta R/R$ gives a picture of the local temperature variations ΔT up to a constant factor of $(1/R)\partial R/\partial T$. Taking an average of images obtained four by four, the signal-to-noise ratio can be increased, whence values of ΔR as small as 10^{-5} can be detected in a few minutes. For longer acquisition times, the signal-to-noise ratio can be still further improved. By establishing a theoretical description of different noise sources and imperfections in the digital-to-analogue conversion of the CCD sensors, R. Ram and coworkers were able to give an analytical expression for the measured signals and an explicit calculation of the error bars [19]. At the expense of a relatively long acquisition time (105 modulation periods, or 4.7 hours for an optimal setup), a record resolution of 18 mK was thereby obtained using a CCD.

Since most cameras operate at maximal frame rates of a few tens of hertz, CCD-based systems are limited to relatively low frequencies. A device in which the illumination is modulated at a different frequency (heterodyning) can be used to work at higher frequencies [17, 20, 21]. Clearly, another way of accessing equivalent information is to work in the time domain. To do this, a pulsed laser is generally used to illuminate the sample. The delay between the electrical excitation and the probe pulse is then adjusted to obtain a measurement of the time dependence of R [22, 23].

13.3 Thermoreflectance Under Visible Illumination

In this wavelength range, most materials used to encapsulate integrated circuits, e.g., Si_3N_4, SiO_2, polymers, alone or combined, are in fact transparent. The light arriving at the device is mainly reflected by the first opaque layer it encounters, e.g., metal or silicon, and the thermal images therefore indicate the temperature of this region [24].

However, the multilayer structure of integrated circuits significantly alters the effective thermo-optical coefficients. Indeed, interference occurs in the transparent passivation layers, modifying both the reflection coefficient R and its temperature dependence as expressed by $\mathrm{d}R/\mathrm{d}T$ [25–28]. A change of a few nanometers in the passivation thickness is enough to change these values radically. The values of $\mathrm{d}R/\mathrm{d}T$ given in the literature for bare materials are then largely irrelevant when these same materials are passivated. Such effects can be exploited to investigate heating in a given material, using one or more suitable wavelengths [29–32].

13.3.1 Spectroscopy of $\mathrm{d}R/\mathrm{d}T$

In order to measure the spectral dependence of the coefficient $\mathrm{d}R/\mathrm{d}T$, a tunable filter or source can be used [22, 23, 32], or white illumination followed by spectroscopic detection. The latter configuration allows faster measurements, since it is multiplexed in wavelength. To do this, the CCD camera is replaced by a CCD spectrometer running as before at 4 times the modulation frequency of the integrated circuit. An alternative is to insert an imaging spectrometer between the microscope and camera, i.e., a system scattering the light from one slit in one direction, while the perpendicular direction conserves spatial information (1D imaging) [28, 31, 33]. In this way, rather than four images per modulation period, one obtains four spectra of reflected intensity variation $I_{\mathrm{therm}}(\lambda)$. It is easy to find the product of the source spectrum and the spectral response of the system, viz., $I_{\mathrm{ref}}(\lambda)$, by using a mirror (e.g., aluminium) of known reflectivity R_{al}, and one can then deduce

$$\left|\Delta R(\lambda)\right| = \left|\frac{\mathrm{d}R}{\mathrm{d}T}(\lambda)\Delta T\right| = \frac{I_{\mathrm{therm}}(\lambda)R_{\mathrm{al}}(\lambda)}{I_{\mathrm{ref}}(\lambda)} \,.$$

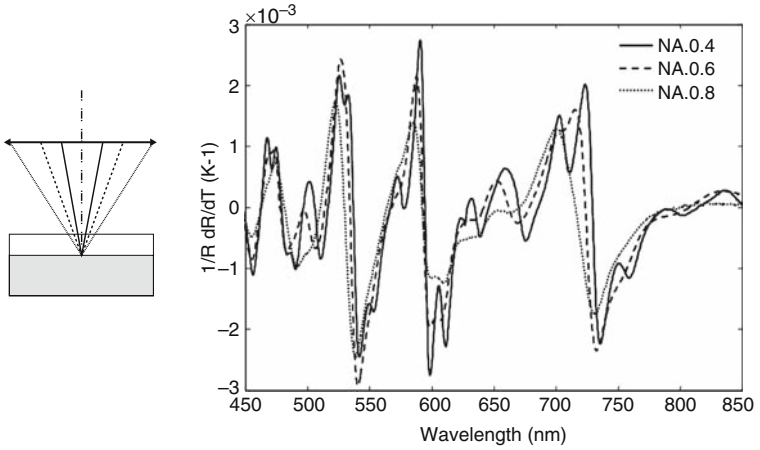

Fig. 13.4 Thermoreflectance spectra $(1/R)\mathrm{d}R/\mathrm{d}T$ measured with 1 min of integration on a polysilicon resistance coated with a layer of Si_3N_4 and polyamide, dissipating a power of 0.12 mW mm^{-2} ($\Delta T_{\mathrm{surf}} = 100°\mathrm{C}$). The three objectives with different numerical apertures collect light in different cones (*left*), which explains the different spectra

To obtain the sign, the phase ϕ of the signal must be considered. The sign of $\mathrm{d}R/\mathrm{d}T$ is then simply the sign of ϕ.

Figure 13.4 gives an example of this type of measurement, obtained in the wavelength range 400–800 nm on a passivated polysilicon integrated resistance, under the same conditions but using objectives with different numerical apertures. The coefficient $\mathrm{d}R/\mathrm{d}T$ varies considerably, and one must of course avoid regions of the spectrum in which it is zero. As can be seen, the resulting spectrum depends significantly on the numerical aperture used. There are many narrow fringes for small numerical apertures, and these broaden for bigger apertures. Indeed, the objective gathers beams over a range of angles of incidence that increases as it becomes more open. For high apertures, spectra from widely different angles of incidence are gathered and averaged in the detector, whence the observed smoothing effect.

13.3.2 Modelling

To interpret the spectra $R(\lambda)$ and $(\mathrm{d}R/\mathrm{d}T)(\lambda)$ in the presence of transparent layers, several kinds of model of have been developed [28, 34, 35]. The simplest approach [28] is to calculate $R(\lambda, T)$ using a matrix formulation, for a temperature T, of the refractive indices $n_i(\lambda, T)$ and thicknesses $e_i(T)$ given for each of the layers (numbered by i). At a temperature $T + \Delta T$, the same model can be used to obtain $R(\lambda, T + \Delta T)$, using the new indices and thicknesses of the materials at $T + \Delta T$:

$$n_i(\lambda, T+\Delta T) \approx n_i(\lambda, T) + \Delta T \frac{\mathrm{d}n_i}{\mathrm{d}T}(\lambda)\,, \qquad e_i(T+\Delta T) = e_i(T)\big(1+\alpha_i \Delta T\big)\,,$$

assuming a linear expansion with coefficient α_i. The coefficient $(\mathrm{d}R/\mathrm{d}T)(\lambda)$ of the multilayer structure is then simply expressed as

$$\frac{\mathrm{d}R}{\mathrm{d}T}(\lambda) \approx \frac{\Delta R(\lambda)}{\Delta T} = \frac{R(\lambda, T+\Delta T) - R(\lambda, T)}{\Delta T}\,.$$

This model is used to calculate $R(\lambda)$ and $(\mathrm{d}R/\mathrm{d}T)(\lambda)$ for a multilayer structure, taking as parameters e_i, α_i, the indices of the materials, and their coefficients $\mathrm{d}n_i/\mathrm{d}T$. However, it is rather simplistic, since the temperature is treated as uniform throughout the structure at any given instant of time.

This approximation may prove to be inadequate if one is concerned with a component of the temperature that is modulated at high frequency, with low diffusion length

$$\mu = \left(\frac{D}{\pi F}\right)^{1/2}\,,$$

where D is the diffusivity. Another analytical model accounting for the temperature distribution was developed by O. Wright et al. [34] for a single transparent layer on an opaque substrate. The temperature profile in the structure is calculated in the 1D approximation, then used to evaluate the indices and thicknesses, and hence also $R(\lambda)$ and $(\mathrm{d}R/\mathrm{d}T)(\lambda)$. A calculation carried out using this model is shown in Fig. 13.5 (left).

However, for higher numbers of layers, exact analytical solution of the problem becomes difficult, and one must resort to numerical methods, dividing the structure up into as many infinitesimal layers as necessary. The heat transfer problem can

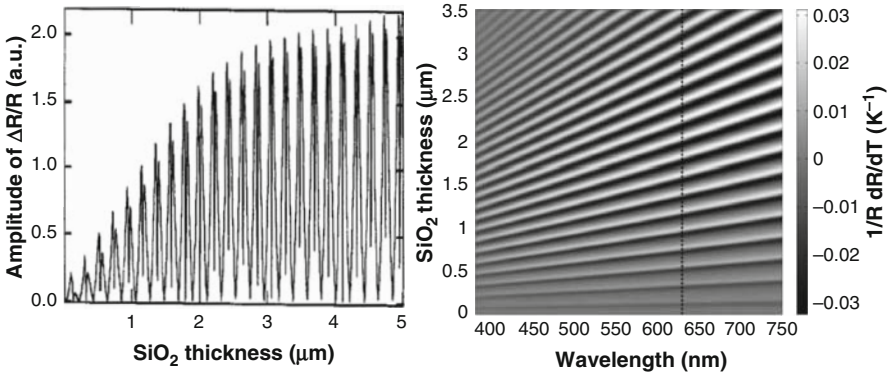

Fig. 13.5 *Left*: Amplitude of $|\Delta R/R|$ as a function of the silica thickness in an SiO_2/Si system, calculated by O. Wright et al. using an analytical model [34] for $\lambda = 632.8$ nm. *Right*: Calculation carried out on the same SiO_2/Si system using a simple model [28]. The amplitude of $(1/R)\mathrm{d}R/\mathrm{d}T$ is represented by a grey scale as a function of the wavelength (*horizontal axis*) and the SiO_2 thickness (*vertical axis*). The two methods predict similar behaviour

be solved for each of these layers. The resulting temperatures can then be used to evaluate the properties of the layers and hence also the optical and thermo-optical characteristics of the whole ensemble [35].

Figure 13.5 shows the results for two of these models on a simple but commonly encountered system in microelectronics, viz., SiO_2/Si. The results are very similar, which suggests that at low frequency, provided that the thermal diffusion length remains shorter than the thickness of the transparent layer, a simple model assuming uniform temperature throughout the system is adequate [28]. In practice, for silica thicknesses typically used in microelectronics ($D = 8.3 \times 10^{-7}$ m^{2}s^{-1}), this limit usually lies between 260 kHz ($\mu_{260\,\mathrm{kHz}} = \sqrt{D/\pi F} = 1$ μm) and 26 MHz ($\mu_{26\,\mathrm{MHz}} =$ 100 nm). Beyond these frequencies, the heat transfer problem must be taken into account [34]. Finally, if the structure contains more than one layer, which is often the case in microelectronics, full numerical solution is required [35].

13.3.3 Calibration

The spectroscopic simulation or measurement methods just presented are essential in order to determine the wavelength required to optimise the coefficient $\mathrm{d}R/\mathrm{d}T$. Furthermore, it is important to know the quantitative value of this coefficient. Given the number of thermal and optical parameters coming into play here, this can only be done experimentally, knowing the temperature of the system.

To obtain a temperature reference, a local probe can be used [30], or if the system is simple enough and the injected power is known, a model can be made to find the temperature [36]. Most current methods [22, 30, 37–43] use a Peltier element to apply a known external temperature excitation to the whole circuit casing. When the probe uses a coherent beam, interference can appear at high magnification between the sample and the objective, requiring accurate focusing control [22, 37, 38]. Such interference does not occur under incoherent illumination, but focusing, or even the lateral positioning of the sample, must nevertheless be controlled owing to motions induced by expansion. This is greatly facilitated by using a camera, since straightforward image analysis can be used to determine the quality of the focusing and the amplitude of lateral displacements.

In the system shown in Fig. 13.6, lateral motions have been simply corrected by calculating the (x,y) shift that optimises the correlation between the given image and a reference image. Vertical motions have been compensated using piezoelectric displacement of the objective, whose position is determined by a 2D gradient calculation (spatial derivative) on the image [39, 44].

In this way, images can be obtained for different temperatures T of the sample, on which the value of the signal $I(T)$ can be recorded in a given region. Since

$$\frac{I(T)}{I(T_0)} = \frac{R(T)}{R(T_0)} \approx 1 + \frac{1}{R(T_0)} \frac{\partial R}{\partial T} (T - T_0) \,,$$

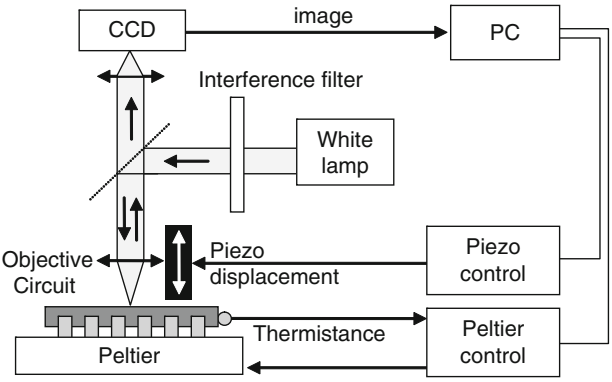

Fig. 13.6 Setup for measuring the calibration coefficient $\mathrm{d}R/\mathrm{d}T$. The temperature of the casing of the integrated circuit is driven by a Peltier element associated with a feedback loop. For each temperature, image analysis determines the optimal focusing and lateral displacement

straightforward linear regression on the values of $I(T)/I(T_0)$ can be used to obtain the slope $[1/R(T_0)]\partial R/\partial T$. If the casing as a whole thermalises slowly, no temperature modulation can be applied. This forbids use of lock-in detection techniques, hence a somewhat mediocre signal-to-noise ratio, but which can be in part compensated by acquiring many points for $I(T)$.

To validate the calibration, we used a specially designed structure with 0.6 mm BCD (Bipolar–CMOS–DMOS, ST Microeletronics) technology. A thermal image of this circuit is shown in Fig. 13.8. It comprises five heating resistances in series, dissipating less than 1 W, and two kinds of probe for making local temperature measurements. On the three central heating resistances, where the most uniform temperature would be expected, a resistive aluminium track of width 1 μm is deposited, with four terminals in the Van der Pauw configuration to eliminate contact resistances: two contacts are used to supply a constant current $1_{\mathrm{VdP}} = 1$ mA to the structure, and the other two to measure the voltage V_{VdP}. This structure was calibrated in an oven for several known temperatures, giving a slope $\mathrm{d}V_{\mathrm{VdP}}/\mathrm{d}T = 1.5 \times 10^{-4}\,\mathrm{V\,K^{-1}}$.

A junction (diode) under a constant voltage $V_{\mathrm{d}} = -1$ V of opposite sign is also included in the setup, between the heating elements. The current in the diode then functions as an extremely accurate and sensitive thermometer. Once again, calibration in an oven gives a slope $\mathrm{d}I_{\mathrm{d}}/\mathrm{d}T = 1.7 \times 10^{-5}\mathrm{A\,K^{-1}}$ for a voltage of -1 V.

The spectra shown previously, in Fig. 13.4, were taken on this structure. As can be seen, the wavelength of $\lambda = 543$ nm is optimal for thermoreflectance measurements if an objective with numerical aperture NA $= 0.6$ is used. For practical reasons, a white lamp was used to illuminate the sample, followed by a band pass filter centered at $\lambda_0 = 536$ nm and of width 13 nm. At this wavelength, the calibration in Fig. 13.7 gives

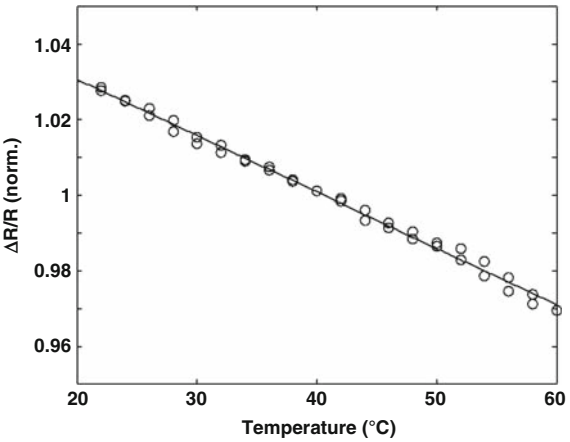

Fig. 13.7 Calibration obtained for $\lambda = 536$ nm (see Fig. 13.4), using a $\times 50$ objective with NA $=$ 0.6 on a passivated polysilicon resistance (Si_3N_4 + polyamide). The *straight line* indicates the best linear fit, with slope -1.49×10^{-3} K^{-1}, and an uncertainty of 0.04×10^{-3} K^{-1}

$$\frac{1}{R(T_0)} \left.\frac{\partial R}{\partial T}\right|_{\text{polySi}} = (-1.49 \pm 0.04) \times 10^{-3}\,\text{K}^{-1}$$

on the passivated polysilicon resistance. Using this coefficient, any thermoreflectance image obtained on polysilicon can thus be converted into a temperature image, provided that the encapsulation thickness and illumination wavelength are the same.

Figure 13.8 shows measurements made on this sample, with square-modulated heating at $F = 7.5$ Hz and for several different peak voltages between 0 and 20 V. The average temperatures measured on the resistance and on the diode are also shown (right-hand graph). The agreement with the measurement delivered by the integrated sensors is excellent, with a standard deviation of only 2.3%. Note that the temperature obtained with the diode is almost 70% less than that of the resistance, owing to the excellent thermal contact between diode and substrate, which contributes significantly to heat dissipation.

This kind of complex structure cannot be modelled analytically, but it can be studied by virtue of computer codes using finite volumes or elements. Figure 13.8 compares a temperature calculation obtained with the TMapper software [45, 46]. The physical parameters used were the specific heat, thermal conductivities, dissipated power, and system geometry, and no fitting was carried out. As expected, the temperature varies linearly with the dissipated power and exhibits good agreement with thermoreflectance measurements. The 3D calculation confirms that the SiO_2 layer separating the resistance from the substrate plays a significant role as thermal insulator and contributes to the high temperatures obtained on the resistance and aluminium thermistances. On the other hand, the region where the diodes are located, between the resistances, is much colder due to the direct thermal contact with the substrate, which has a thermal conductivity almost 100 times greater than SiO_2.

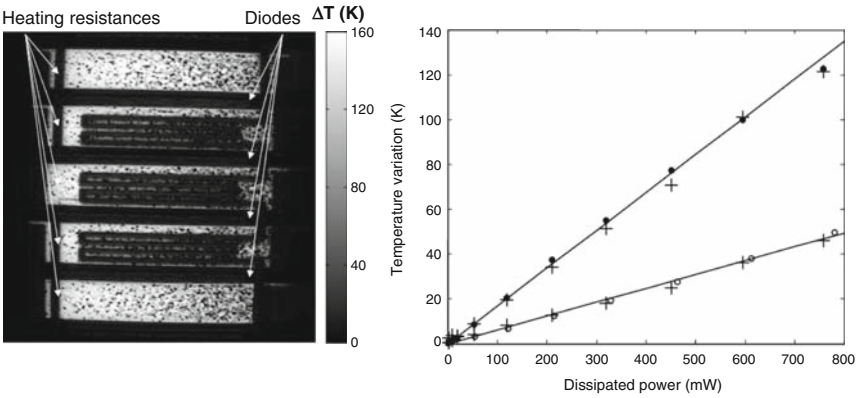

Fig. 13.8 *Left*: Thermal images (90×90 μm) obtained on integrated polysilicon resistances subjected to square modulation with $F = 7.5$ Hz and $V_{pp} = 20$ V, for $\lambda_0 = 536$ nm. The temperature scale is only valid on the polysilicon resistance. Acquisition time 1 min. *Right*: Average temperature measured by thermoreflectance (*crosses*) on the resistance and above the diode, for different heating powers. *Black circles* represent the temperatures obtained with the 4-point aluminium thermistance and *white circles* the temperature given by the diode. *Continuous lines* represent TMapper simulations

13.4 Thermoreflectance Under Ultraviolet Illumination

As we have just seen, the coefficient $\mathrm{d}R/\mathrm{d}T$ depends heavily on the kind of materials and the interference occurring in the passivation layers, which can considerably complicate the problem of obtaining quantitative values for the temperature in structures comprising several materials. It is nevertheless possible to exploit the high absorption by passivation dielectrics in the ultraviolet (UV) part of the spectrum. On these opaque layers, the reflectivity can be measured, as on any other material, in such a way as to obtain a temperature map at the passivation surface, independently of the underlying materials. The information obtained is then close to that delivered by scanning local probes [47].

Si_3N_4 does not absorb significantly above 280 nm, but this absorption then grows monotonically in the deep UV. Most compounds based on Si_3N_4 reach absorption levels in the range 5×10^6 to 2×10^7 m^{-1}, even below the absorption threshold for UV by oxygen in the air, around 200 nm [11]. Without needing to work in vacuum, absorptions and reflectivities comparable to those of metals can thus be obtained on Si_3N_4 dielectric surfaces. In practice, these layers become sufficiently absorbent to screen the underlying materials for $\lambda < 250$ nm. Interference effects are then eliminated, and a single, uniform material is visible at the surface. It therefore suffices to know its coefficient $\mathrm{d}R/\mathrm{d}T$ in order to obtain a quantitative temperature at all points of the surface [12].

In the case presented below to illustrate this technique, measurements were carried out using a single-beam setup with lock-in detection rather than a camera. The source is then a frequency-tripled Ti:sapphire laser ($\lambda = 720$ nm) at $\lambda/3 = 240$ nm.

Note, however, that the coherence of the laser was not needed here, and that these experiments could be carried out with any UV source. The rate of this pulsed laser, of the order of 100 MHz, is well above the frequencies used for this experiment and exceeds the pass band of the electronics used, whence the beam may be considered continuous. The beam is filtered to eliminate any residues at λ and $\lambda/2$, then finely focused using a fused silica lens. The polarisation is perpendicular to the plane of incidence. After reflection, the light passes through a narrow band pass filter centered on 240 nm to eliminate unwanted light, then measured with a photodiode.

The electronic circuit used is a simple series of NiCr (154 ohm) resistances deposited on a GaAs substrate, supplied by gold tracks and coated with Si_3N_4 [30]. Each resistance receives a square modulation 0–10 V with frequency between 0.2 and 1 600 Hz and dissipates an average power of 325 mW. Images are then obtained by scanning the sample in front of the beam. The resolution is given by the size of the focusing spot, viz., 1 μ in the case presented here, but in an optimised setup, a resolution of the order of $0.6\lambda/\mathrm{NA} \approx 240$ nm can be expected for a numerical aperture of $\mathrm{NA} = 0.6$, which is typical for UV objectives.

Under UV illumination, as can be seen from Fig. 13.9, the heat source and diffusion in the substrate are clearly visible. The profile on the right of the image is remarkably continuous, unlike what is observed in the visible [12, 30], which gives a good indication of the consistency of the result. To obtain quantitative temperature values, calibration was carried out by a similar method to the one discussed in the last section, with the result $\mathrm{d}R/\mathrm{d}T \approx \Delta R/\Delta T = (3 \pm 0.5) \times 10^{-5}\ \mathrm{K}^{-1}$.

Temperatures measured with this technique are mainly those of the upper surface. Diffusion through the passivation layer significantly spreads the heat, and this reduces the spatial resolution [48]. At low frequencies, the passivation layer can nevertheless be considered as thermally thin. In this experiment, with a thermal diffusivity $D_{\mathrm{Si_3N_4}} = 0.23\ \mathrm{cm^2 s^{-1}}$, and at $F = 1\,600$ Hz, the thermal diffusion length $\mu = \sqrt{D/\pi F}$ is 60 μm, i.e., two orders of magnitude longer than the thickness of the passivation layer. However, as mentioned above, a surface technique like this cannot be used to image phenomena shorter than 50 ns or at frequencies higher than 20 MHz.

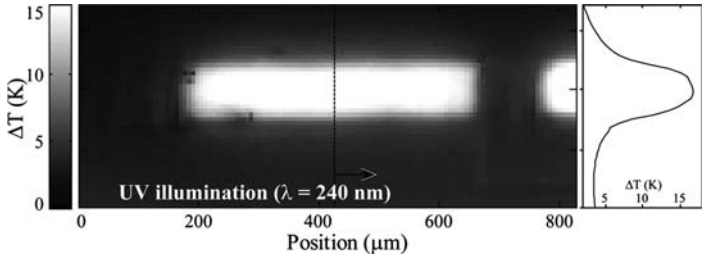

Fig. 13.9 Thermal image of an integrated resistance (NiCr on GaAs) under UV illumination at $\lambda = 240$ nm. The temperature profile *on the right* is perfectly continuous, in contrast to what is observed under visible illumination, when the various materials are visible under the passivation layer

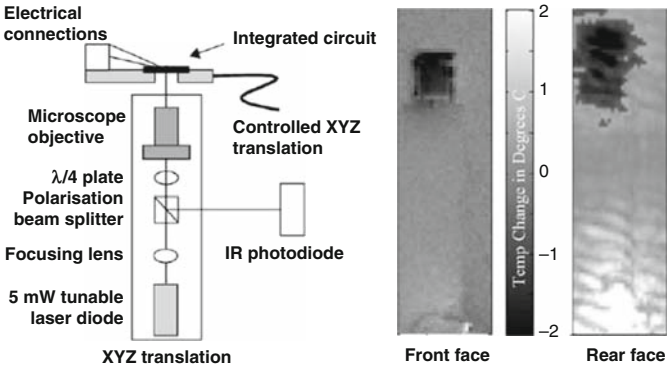

Fig. 13.10 Near-infrared single beam setup and front and rear face images of a microscale Peltier element. The rear face image was obtained through 200 μm of silicon, using the maxima of images made at four different wavelengths in order to eliminate interference effects in the substrate. From [51]

13.5 Thermoreflectance in the Near Infrared. Rear Face Imaging

The silicon used in the vast majority of integrated circuits is perfectly transparent at wavelengths greater than 1.1 μm. These wavelengths thus constitute an excellent means for observing integrated circuits through their substrate [49, 50], but also to obtain thermoreflectance measurements. The light reflected by the interface between the silicon and the other materials then provides information about the active layers [51, 52].

13.5.1 Near-Infrared Thermoreflectance with Laser Illumination

To carry out this type of measurement, a tunable laser can be used in the near-infrared, focused through the silicon substrate onto the opposite, active, face of the device [51]. The modulation of the reflected light is measured, as in a conventional thermoreflectance setup, using lock-in detection. The main difficulty lies in the fact that the coherence length of the laser is greater than the thickness of the silicon substrate (200 μm here). Interference results, and this makes precise measurements of the reflectivity variations somewhat difficult. By varying the illumination wavelength, it is nevertheless possible to shift these interference fringes. The image in Fig. 13.10 shows the maxima of the thermoreflectance images acquired at different wavelengths.

Another difficulty lies in the high reflectivity of silicon. A large part of the light received by the detector comes from specular reflection on the rear face of the substrate, which dazzles the detector while being of little interest from the standpoint of thermal analysis. Polishing and antireflection treatment of the rear face are a

possibility but tedious to implement. Rampnoux et al. have succeeded in eliminating this reflection by using time gating [52], a technique originally developed for imaging applications in scattering media [53]. A femtosecond laser is split into two beams. The first illuminates the circuit, reflecting on the rear face of the substrate and on the internal active face of the integrated device, before being combined in a type 2 frequency-doubling crystal with the second, reference, beam. Given the short duration of the light pulses, one only has to choose the path travelled by the reference so that a pulse arrives at the same time as the pulse reflected on the active face. In the crystal, only those pulses that arrive simultaneously will give rise to second harmonic generation. The visible image produced by these two infrared pulses comes exclusively from the active interface of the circuit. However, this elegant method also suffers from interference occurring in the substrate.

13.5.2 Near-Infrared Thermoreflectance with Incoherent Illumination

To eliminate this phenomenon, it is interesting to use a broad spectrum illumination, with coherence length less than the substrate thickness [54]. To do this, a conventional filament lamp can be used, which emits strongly in the near-infrared. The setup with a microscope and camera described previously can be used, but remembering to remove the filters inserted in the microscope to eliminate this generally undesirable infrared radiation. We used an InGaAs Sensors Unlimited camera (SU320 MS, 12 bits, quantum efficiency > 0.6 in the range 1–1.7 μm) running at a frame rate of $4F = 20$ Hz, for an integration time of 8 ms. The circuit is supplied with a square-modulated voltage at $F = 5$ Hz, inducing heating and reflectivity variations at the same frequency.

The measurements shown here were obtained on an aluminium track of width 3 μm and length 200 μm deposited on a titanium coupling layer (20 nm), with Au contacts of side 100 nm. This structure was made on an undoped silicon substrate of thickness 500 μm, polished on both faces. This 29 ohm resistance, supplied between 0 and 4.2 V and 145 mA, dissipated a peak power of 609 mW.

The image shown in Fig. 13.11 can deliver quantitative information if the thermo-optical coefficients of each material are known. In the silicon substrate, one can use $\mathrm{d}n/\mathrm{d}T = 1.835 \times 10^{-4}\ \mathrm{K}^{-1}$ [55], whence

$$\left.\frac{1}{R}\frac{\partial R}{\partial T}\right|_{\mathrm{Si}} = \frac{4}{(n-1)(n+1)}\frac{\mathrm{d}n}{\mathrm{d}T} = 6.6 \times 10^{-5}\ \mathrm{K}^{-1}\,.$$

We may deduce that, close to the resistance, where $\Delta R/R \approx 1.8 \times 10^{-3}$ (see Fig. 13.11), there is a temperature rise $\Delta T_{\mathrm{Si}} \approx 27$ K. These measurements clearly show that no interference occurs in the substrate when incoherent illumination is used. Moreover, unwanted reflection at the air/silicon interface is not modulated and is therefore eliminated by the lock-in detection.

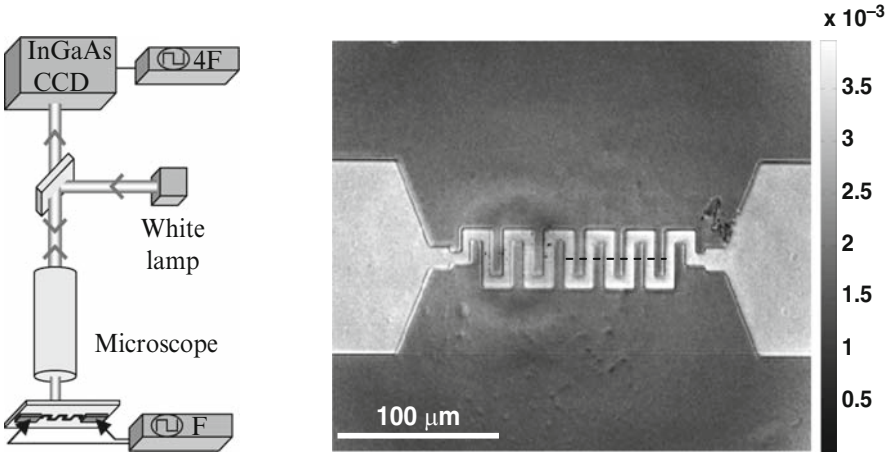

Fig. 13.11 *Left*: Near-IR thermoreflectance setup for imaging through silicon. The IC is supplied by connections on the rear face. *Right*: Thermoreflectance image of a resistance dissipating 609 mW (peak) obtained through 500 μm of silicon with a ×50 objective, NA = 0.6. The resolution measured along the *dashed line* is 1.7 μm. The temperature in the immediate vicinity of the resistance is $\Delta T_{\mathrm{Si}} \approx 27$ K

The spatial resolution of the images obtained with all the methods just discussed is nevertheless rather limited. Whether the beam is focused or not, when using a time gate, one cannot hope for a better resolution than 670 nm for wavelengths greater than 1.1 μm, even if one had a hypothetical numerical aperture of 1. For example, the resolution of the image in Fig. 13.11, estimated by deriving the profile along the dashed line, is 1.7 μm. Given that the camera is sensitive up to $\lambda = 1.7$ μm and that the objective has numerical aperture NA = 0.6, this corresponds to the diffraction limit $1.22 \times (\lambda/2) \times \mathrm{NA} = 1.7$ μm. For many applications in integrated electronics, such resolutions are not good enough.

13.5.3 Improving Resolution with a Solid Immersion Lens

In conventional microscopy, the diffraction limit can be slightly improved by reducing λ or by increasing the numerical aperture NA. Since imaging through silicon requires one to work in the transparency range $\lambda > 1.1$ μm, the only free parameter is NA. Immersion in water or oil provides a way of exceeding the limit NA = 1. In solids, the same can be achieved by using plane–convex lenses with the same index as the substrate [56], called solid immersion lenses (SIL), or numerical aperture increasing lenses (NAIL). Recently, these devices have been adapted for application to silicon [49, 57, 58]. They can obtain a stigmatic image provided that the relation

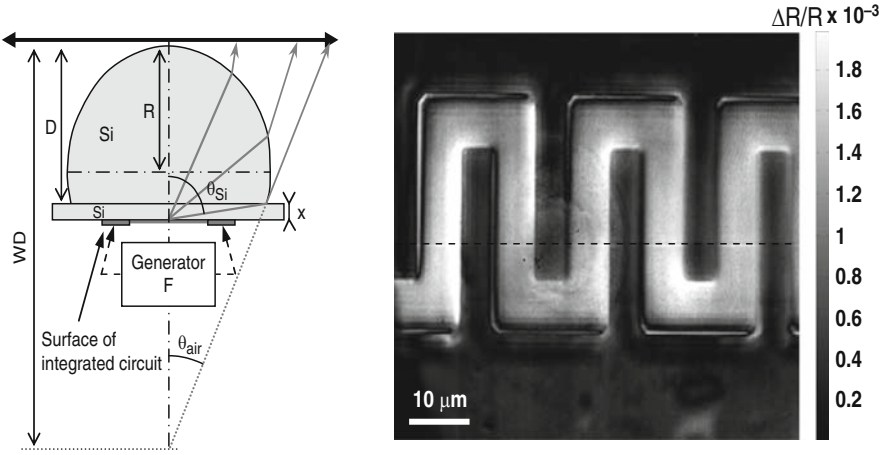

Fig. 13.12 *Left*: Schematic view of a solid immersion lens with $X = 500$ μm, $R = 3.5$ mm, and $D = 4$ mm. *Right*: Thermal image of an aluminium resistance dissipating 609 mW, obtained through a silicon substrate of thickness 500 μm and a SIL with a ×50 objective and NA = 0.6. The resolution measured along the *dashed line* is 440 nm, corresponding to an effective aperture value of 2.36

$$D + X = R\left(1 + \frac{1}{n_{\mathrm{Si}}}\right)$$

is satisfied (see Fig. 13.12 for the notation). The light collection angle is then significantly increased due to the very high refractive index of silicon, viz., $n_{\mathrm{Si}} = 3.5$, in the near infrared [11]. In theory it is possible to reach numerical apertures of 3.5 rather than 1 in air, which corresponds to an improvement in the resolution by a factor of 3.5. In the field of integrated circuit inspection, configurations of this kind have been used for reflection imaging on the rear face [49], two-photon absorption imaging [50], and thermal emission imaging in the range 8–12 μm [59, 60]. In the present case, using the near infrared up to $\lambda = 1.1$ μm, a theoretical resolution as low as $1.22\lambda/2n_{\mathrm{Si}} = 192$ nm can be expected.

In the present case [54], for a substrate of thickness $X = 500$ μm, the LIS is aplanatic, i.e., $D > R$, with $R = 3.5$ m and $D = 4$ mm. A simple calculation of geometrical optics yields the maximal light gathering angle in the silicon as $\theta_{\mathrm{Si}} = 81.8^\circ$, corresponding to $\mathrm{NA}_{\mathrm{SIL}} = n_{\mathrm{Si}} \sin\theta_{\mathrm{Si}} = 3.47$, close to the limit of 3.5. These beams emerge from the lens with an angle $\theta_{\mathrm{air}} = 16.4^\circ$ corresponding to $\mathrm{NA}_{\mathrm{air}} = \sin\theta_{\mathrm{air}} = 0.28$, and are thus easily picked up by most microscope objectives. Once these rays are correctly collected, increasing the aperture of the microscope objective beyond this figure will not improve the resolution. The main difficulty then comes from the large working distance WD between the virtual object plane, located well below the surface of the integrated circuit, and the upper face of the SIL. To be able to focus, the objective must therefore have a longer working distance than 1.58 cm. In this case, a ×20 Nikon objective with NA = 0.33 was used, allowing a long working distance WD = 0.6.

The result shown in Fig. 13.12 was achieved under the same conditions as before ($F = 5$ Hz, peak power 609 mW). The thermal image shows stronger heating near the corners of the resistive track, as expected in these regions where the electron density is higher. The resolution is estimated at 440 nm, once again by deriving a profile (along the dashed line), and by measuring the full-width at half-maximum of the resulting peaks. This resolution corresponds to an effective numerical aperture $NA_{SIL} = 2.36$ at $\lambda = 1.7$ μm, certainly inside the theoretical prediction, but much better than the diffraction limit in air, which would be 3.1 μm with the objective of numerical aperture 0.33 used here, and 1 μm with an ideal objective of numerical aperture 1.

Quantitatively, the signals obtained with the SIL are $\Delta R/R = 5.2 \times 10^{-4}$ in Si and $\Delta R/R = 2.2 \times 10^{-3}$ on the resistance. Although the electrical conditions are the same, the values are respectively 2.8 and 1.7 times lower than in the last experiment, without the SIL. By acting as a dissipator, the silicon lens may contribute to this effect, but finite element simulations have shown that it has only a slight influence. Since the temperatures are probably the same in the two experiments, this difference must very likely be attributed to a difference between the coefficients dR/dT of the two configurations. As shown above in the case of visible illumination (see Fig. 13.4), these coefficients depend heavily on the numerical aperture used, and are thus likely to differ with NA = 0.6 (without SIL, $\theta = 36°$) and NA = 2.36 (with SIL, $\theta_{Si} = 81.8°$). This very likely explains the differences obtained with the two techniques, while emphasising the need for specific calibration in each new measurement configuration.

All the techniques discussed here may be able to meet the requirements for thermal imaging of the rear face, particularly in those cases where the active zones are coated with opaque layers. In each case, the rear face must be polished in order to minimise aberration, but this raises no particular difficulty in most production processes. The presence of doped silicon, which may be strongly absorbing in the near infrared, can nevertheless prove problematic. In this case, the temperatures measured will be those of the doped zones which will reflect the light. The ultimate resolution limit with SILs should be attainable by optimising the imaging arrangements, and in particular, by reducing chromatic aberration. To do this, it will be useful to illuminate with a narrower band centered on 1.1 μm. At a resolution of about 200 nm, and given current integration densities, these techniques should then provide a way of studying thermal phenomena in most integrated circuits.

References

1. A. Rosencwaig and A. Gersho: Journal of Applied Physics **47**, 64 (1976)
2. A. Rosencwaig, J. Opsal, W.L. Smith, et al.: Applied Physics Letters **46**, 1013 (1985)
3. D.Y. Li, Y.Y. Wu, P. Kim, et al.: Applied Physics Letters **83**, 2934 (2003)
4. P. Kim, L. Shi, A. Majumdar, et al.: Physical Review Letters **8721** (2001)
5. G. Abstreiter: Applied Surface Science **50**, 73 (1991)
6. M. Cazzanelli, C. Vinegoni, and L. Pavesi: Journal of Applied Physics **85**, 1760 (1999)

7. L. Pavesi and M. Guzzi: Journal of Applied Physics **75**, 4779 (1994)
8. S. Grauby, A. Salhi, J.M. Rampnoux, et al.: Review of Scientific Instruments **78**, 74902 (2007)
9. S. Dilhaire, D. Fournier, and G. Tessier: Microscale and Nanoscale Heat Transfer **107**, 239 (2007)
10. G. Ghosh: *Handbook of Thermo-Optic Coefficients of Optical Materials with Applications*, Academic Press, New York (1998)
11. E.D. Palik: Journal of the Optical Society of America, Optics Image Science and Vision **1**, 1297 (1984)
12. G. Tessier, S. Hole, and D. Fournier: Optics Letters **28**, 875 (2003)
13. W.J. Scouler: Physical Review Letters **18**, 445 (1967)
14. R. Rosei and D.W. Lynch: Physical Review B **5**, 3883 (1972)
15. J. Christofferson and A. Shakouri: Review of Scientific Instruments **76** (2005)
16. A. Shakouri: Proceedings of the IEEE **94**, 1613 (2006)
17. S. Grauby, B.C. Forget, S. Hole, et al.: Review of Scientific Instruments **70**, 3603 (1999)
18. P. Gleyzes, F. Guernet, and A.C. Boccara: Journal of Optics Nouvelle Revue D Optique **26**, 251 (1995)
19. P.M. Mayer, D. Luerssen, R.J. Ram, et al.: Journal of the Optical Society of America, Optics Image Science and Vision **24**, 1156 (2007)
20. C. Filloy-Corbrion, G. Tessier, S. Hole, et al.: Journal de Physique IV **125**, 105 (2005)
21. B.C. Forget, S. Grauby, D. Fournier, et al.: Electronics Letters **33**, 1688 (1997)
22. M.G. Burzo, P.L. Komarov, and P.E. Raad: IEEE Transactions on Components and Packaging Technologies **28**, 637 (2005)
23. M.G. Burzo, P.L. Komarov, and P.E. Raad: IEEE Transactions on Components and Packaging Technologies **28**, 39 (2005)
24. G. Tessier, C. Filloy, M.L. Polignano, et al.: Journal de Physique IV **125**, 423 (2005)
25. R. Abid and F. Miserey: Comptes Rendus de l'Academie des Sciences Serie II **319**, 631 (1994)
26. R. Abid, F. Miserey, and F.Z. Mezroua: Journal de Physique III **6**, 279 (1996)
27. V. Quintard, G. Deboy, S. Dilhaire, et al.: Microelectronic Engineering **31**, 291 (1996)
28. G. Tessier, G. Jerosolimski, S. Hole, et al.: Review of Scientific Instruments **74**, 495 (2003)
29. S. Hole, G. Tessier, C. Filloy, et al.: Electronics Letters **38**, 986 (2002)
30. G. Tessier, S. Hole, and D. Fournier: Applied Physics Letters **78**, 2267 (2001)
31. G. Bautista, C. M. Blanca, S. Delica, et al.: Optics Express **14**, 1021 (2006)
32. L.R. de Freitas, E.C. da Silva, A.M. Mansanares, et al.: Journal of Applied Physics **98**, 63508 (2005)
33. G. Bautista, C.M. Blanca, and C. Saloma: Applied Optics **46**, 855 (2007)
34. O.B. Wright, R. Li Voti, O. Matsuda, et al.: Journal of Applied Physics **91**, 5002 (2002)
35. G. Jerosolimski, V. Reita, G. Tessier, et al.: Journal de Physique IV **125**, 419 (2005)
36. S. Dilhaire, S. Jorez, L.D. Patiño Lopez, et al.: In: *Heat Transfer and Transport Phenomena in Microsystems*, Banff, Canada, (2000), pp. 392–397
37. S. Dilhaire, S. Grauby, and W. Claeys: Applied Physics Letters **84**, 822 (2004)
38. S. Dilhaire, S. Grauby, and W. Claeys: IEEE Electron Device Letters **26**, 461 (2005)
39. G. Tessier, M.L. Polignano, S. Pavageau, et al.: Journal of Physics D Applied Physics **39**, 4159 (2006)
40. V.V. Pavlov, G. Tessier, C. Malouin, et al.: Applied Physics Letters **75**, 190 (1999)
41. G. Tessier and P. Beauvillain: Applied Surface Science **164**, 175 (2000)
42. V.V. Pavlov, J. Ferre, P. Meyer, et al.: Journal of Physics Condensed Matter **13**, 9867 (2001)
43. G. Tessier, A. Salhi, Y. Rouillard, et al.: Journal de Physique IV **125**, 375 (2005)
44. G. Tessier, S. Pavageau, B. Charlot, et al.: IEEE Transactions on Components and Packaging Technologies **30**, 604 (2007)
45. J.S. Wilson and P.E. Raad: International Journal of Heat and Mass Transfer **47**, 3707 (2004)
46. P.E. Raad, P.L. Komarov, and M.G. Burzo: IEEE Transactions on Components and Packaging Technologies **30**, 597 (2007)
47. L. Aigouy, G. Tessier, M. Mortier, et al.: Applied Physics Letters **87** (2005)

48. S. Gomes, P.O. Chapuis, E. Nepveu, et al.: IEEE Transactions on Components and Packaging Technologies **30**, 424 (2007)
49. S.B. Ippolito, B.B. Goldberg, and M.S. Unlu: Applied Physics Letters **78**, 4071 (2001)
50. E. Ramsay, N. Pleynet, D. Xiao, et al.: Optics Letters **30**, 26 (2005)
51. J. Christofferson and A. Shakouri: Microelectronics Journal **35**, 791 (2004)
52. J.M. Rampnoux, H. Michel, M.A. Salhi, et al.: Microelectronics Reliability **46**, 1520 (2006)
53. C. Doule, T. Lepine, P. Georges, et al.: Optics Letters **25**, 353 (2000)
54. G. Tessier, M. Bardoux, C. Boue, et al.: Applied Physics Letters **90** (2007)
55. H.H. Li: Journal of Physical and Chemical Reference Data **9**, 561 (1980)
56. S.M. Mansfield and G.S. Kino: Applied Physics Letters **57**, 2615 (1990)
57. S.B. Ippolito, B.B. Goldberg, and M.S. Unlu: Journal of Applied Physics **97** (2005)
58. E. Ramsay, K.A. Serrels, M.J. Thomson, et al.: Applied Physics Letters **90** (2007)
59. S.B. Ippolito, S.A. Thorne, M.G. Eraslan, et al.: Applied Physics Letters **84**, 4529 (2004)
60. O. Breitenstein, F. Altmann, T. Riediger, et al.: Microelectronics Reliability **46**, 1508 (2006)

Chapter 14
Thermal Microscopy with Photomultipliers and UV to IR Cameras

Bernard Cretin and Benjamin Rémy

As techniques have evolved, there has been an increasing need for better tools for process control and product testing. In the area of thermal measurements, as in other fields such as microscopy, telemetry, and so on, optical methods have often provided measurement solutions, by their non-invasive nature. Indeed, the heat sink phenomenon [1], which occurs whenever a material sensor is placed close to the object whose temperature is to be measured, as happens with near-field techniques (SThM methods [2] or thermal AFM where the tip is either in direct contact or very close to the surface), does not arise in optics, where the measurement can be considered non-perturbing.

Various optical methods have been tested for temperature measurements: radiometry, historically the first [3, 4], thermoreflectance [5, 6] described in Chap. 13, and fluorescence [7–9]. Recently, near-field or SNOM techniques have been able to obtain excellent spatial resolution [10–12]. The main advantage with the latter techniques lies in their high spatial resolution. In contrast to far-field radiometric methods, this resolution is no longer limited by diffraction (Rayleigh criterion, where the ultimate spatial resolution is of the order of half the wavelength $\Delta x \approx \lambda/2$). However, it also depends sensitively on the geometry of the probe and the probe–sample distance, and it is therefore difficult to control, all the more so as the processes used to make probes remain poorly reproducible today. Furthermore, owing to the close proximity of the probe and surface, it necessarily perturbs the temperature of the surface, even if there is no real physical 'contact' between probe and surface. This proximity forbids any high temperature measurement, which would damage or destroy the probe [13]. Finally, it is not easy to interpret and analyse the measured signal, which contains not only thermal but also topological information.

In this chapter, we shall be concerned only with short wavelength radiometric techniques, which provide an excellent compromise between sensitivity, spatial resolution, and measurement accuracy. Indeed, conventional radiometry uses long wavelengths, typically in the range 5–14 μm, depending on the temperature to be measured. This therefore tends to limit the resolution, and involves two associated drawbacks, namely, lower resolution and greater sensitivity to the emissivity of the

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 411
DOI 10.1007/978-3-642-04258-4_14, © Springer-Verlag Berlin Heidelberg 2009

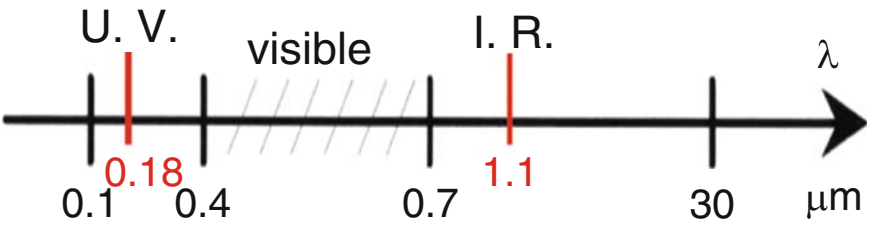

Fig. 14.1 Limits of the various radiometric techniques

surface. In Sect. 14.1, we review the physics underlying the basic relations and the way they are applied to specific instruments. In Sect. 14.2, we describe the components of the measurement instruments and the various kinds of experimental setup that can be used for short wavelength radiometry, from the photon-counting photomultiplier to the ultraviolet or infrared camera. A physical and technical analysis of the different setups will lead to a definition of their main features, the reasons for choosing them, and a discussion of their particular limitations. We shall also consider the problems of signal and image processing, essential for calibrating the system and obtaining good temperature resolution. We shall then explain how the measurement limitations can be pushed back, with some examples of applications demonstrating the quality of the results on different scales, and a brief discussion of the prospects for future developments and applications, especially in microtechnology. The targeted wavelength range is shown in Fig. 14.1.

14.1 Basic Physics

14.1.1 Radiometry

In this section we review the key notions required to understand radiometric measurement. The reader wishing to find out more about the physics and techniques of photon sensors is referred to the literature [4, 14–17].

Flux

The flux ϕ is the instantaneous value of a radiative flow (conservative if propagating in a uniform and non-absorbing medium). It is a power, given in watts (W). In practice, a power measurer gives the value of the flux emitted by a source.

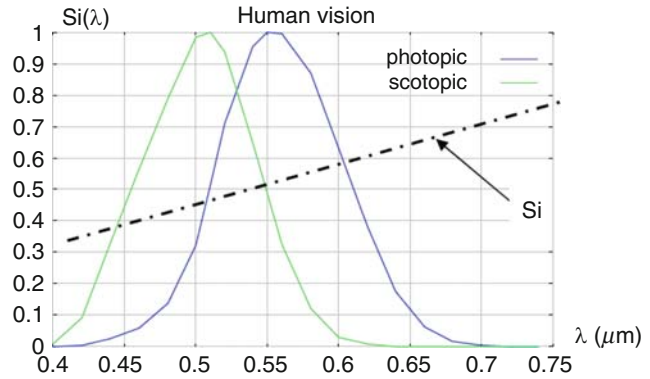

Fig. 14.2 Sensitivity of the human eye compared with the ideal sensitivity of silicon (Si)

Sensitivity. Comparison with Ocular Perception

Like any quantum photodetector, the eye is a non-linear detector, i.e., its response depends on the illumination and not on the electric field. For low levels of illumination, the sensitive elements in the retina are the rod cells (scotopic vision), while at high levels, it is the cone cells that take the leading role (photopic vision). Figure 14.2 shows how chromatic sensitivity $S(\lambda)$ depends on the lighting conditions.

Beam Throughput

Figure 14.3 recalls the definition of solid angle. $\Omega = 1$ steradian if the area covered on a sphere of unit radius is equal to 1:

$$\Omega = \frac{S}{d^2} \,. \tag{14.1}$$

So for 1 steradian, the area covered on a sphere of radius R will be R^2. Since the area of a sphere is $4\pi R^2$, the solid angle describing the sphere is 4π.

Considering Fig. 14.4, $\mathbf{N}$ is the vector normal to the surface and $\mathrm{d}\Omega$ the solid angle subtended by one of the surface elements at the other. θ is the angle of emergence (S) or incidence (R) and d the distance between $\mathrm{d}S$ and $\mathrm{d}R$. Then

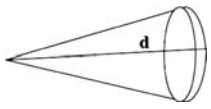

Fig. 14.3 Definition of solid angle

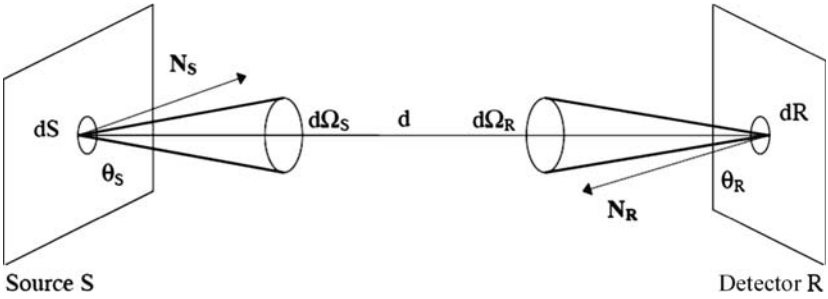

Fig. 14.4 Illustrating the notion of throughput

$$\mathrm{d}\Omega_\mathrm{S} = \frac{\mathrm{d}R\cos\theta_\mathrm{R}}{d^2}\,, \qquad \mathrm{d}\Omega_\mathrm{R} = \frac{\mathrm{d}S\cos\theta_\mathrm{S}}{d^2}\,. \tag{14.2}$$

The elementary throughput of the pencil subtended by dS or dR is defined by

$$\mathrm{d}^2G = \mathrm{d}S\,\mathrm{d}\Omega_\mathrm{S}\cos\theta_\mathrm{S} = \mathrm{d}R\,\mathrm{d}\Omega_\mathrm{R}\cos\theta_\mathrm{R} = \frac{\mathrm{d}S\,\mathrm{d}R\cos\theta_\mathrm{S}\cos\theta_\mathrm{R}}{d^2}\,. \tag{14.3}$$

The total throughput is then expressed in the form

$$G = \int_\mathrm{S}\int_\mathrm{R} \mathrm{d}^2G\,, \tag{14.4}$$

with units of $\mathrm{m}^2\mathrm{sr}$.

Luminance

We assume that the solid angles $\mathrm{d}\Omega_\mathrm{S}$ and $\mathrm{d}\Omega_\mathrm{R}$ are small and that the flux is uniform. The luminance of the radiation emitted by dS and received by dR is

$$L_\mathrm{S} = \frac{\mathrm{d}^2\Phi_\mathrm{S}}{\mathrm{d}^2G}\,, \tag{14.5}$$

where $\mathrm{d}^2\Phi_\mathrm{S}$ is the flux element. The luminance has units of $\mathrm{W/m}^2\mathrm{sr}$.

Illumination

This is the flux per unit area arriving at a surface, given by

$$E = \frac{\mathrm{d}\Phi_\mathrm{R}}{\mathrm{d}R} = \int_{\text{half space}} L_\mathrm{R}\cos\theta_\mathrm{R}\,\mathrm{d}\Omega_\mathrm{R}\,, \tag{14.6}$$

in units of $\mathrm{W/m}^2$.

Emittance or Exitance

This is the flux per unit area leaving the surface of a source, given by

$$M = \frac{\mathrm{d}\Phi_\mathrm{S}}{\mathrm{d}S} = \int_{\text{half space}} L_\mathrm{S} \cos\theta_\mathrm{S} \mathrm{d}\Omega_\mathrm{S} \,, \tag{14.7}$$

also in units of W/m^2.

Luminous Intensity

For sources that are not omnidirectional, this is defined by the relation

$$I = \frac{\mathrm{d}\Phi_\mathrm{S}}{\mathrm{d}\Omega_\mathrm{S}} = \int_\mathrm{S} L_\mathrm{S} \cos\theta_\mathrm{S} \mathrm{d}S \,. \tag{14.8}$$

In general, I depends on the direction angles ξ and η. The surface traced out by the end of $I(\xi, \eta)$ in spherical coordinates is the intensity indicator of the source.

Exposure

This is the energy received per unit area, i.e., the integral of the illumination over the exposure time, expressed by

$$H = \int_{t_1}^{t_2} E(t) \mathrm{d}t \,, \tag{14.9}$$

and given in units of J/m^2. Note that this relation is useful in imaging with a camera or CCD camera, and particularly in the situations to be described in this chapter, since the number of photons received in the photodetector is proportional to the measurement time.

14.1.2 Black Body Emission and Planck's Law

Here we consider an incoherent source obtained by heating an emissive body, e.g., radiator, incandescent lamp. The emission from a black body, i.e., an imaginary object able to absorb all radiation, is given by the Planck relation

$$M^0_\lambda(T) = \frac{\mathrm{d}M_\mathrm{BB}}{\mathrm{d}\lambda} = \frac{2\pi hc^2 \lambda^{-5}}{\exp(hc/\lambda kT) - 1} = \frac{C_1 \lambda^{-5}}{\exp(C_2/\lambda T) - 1} \,, \tag{14.10}$$

where $h = 6.62 \times 10^{-34}$ J s is Planck's constant, $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and c is the speed of light. The constants C_1 and C_2 are

$$C_1 = 2\pi hc^2 = 3.741 \times 10^{-16}\,\mathrm{W\,m^2}\,, \qquad C_2 = \frac{hc}{k} = 0.01488\,\mathrm{m\,K}\,.$$

The quantity $M^0_\lambda(T)$ is expressed in W/m^3.

Figure 14.5a shows the monochromatic emittance of a black body for several different temperatures. The emission maximum is observed to move to shorter wavelengths as the temperature increases. The wavelength $\lambda_{\max}$ corresponding to the maximum of the Planck distribution is given by Wien's law (value of the wavelength where the derivative of the Planck function vanishes):

$$\lambda_{\max} = \frac{2898}{T}\,, \tag{14.11}$$

expressed in μm. In this case,

$$M^0_{\lambda_{\max}}(T) \approx 1.286 \times 10^{-5} T^5\,. \tag{14.12}$$

This function is plotted in Fig. 14.5b. It justifies the use of a specific wavelength range depending on the temperature of the body under investigation. For example, for a temperature of 300 K, the emission maximum corresponds approximately to a wavelength of 10 μm. In the case of an incandescent lamp, whose filament typically heats up to a temperature of 2 700 K, the same relation shows that the emission maximum lies in the near infrared and that the efficiency of the lamp can only be very low (5–8%). Note that most of the thermal emission occurs for $\lambda > \lambda_{\max}$ (roughly 75% of the total).

In practice, typical bodies often behave quite differently to black bodies, especially at short wavelengths where electron resonances begin to occur. To account for the optical properties of surfaces, we define the spectral emissivity $\varepsilon(\lambda)$, denoted by ε_λ in what follows:

$$\varepsilon(\lambda) = \frac{M_\lambda(T)}{M^0_\lambda(T)} \quad \Longrightarrow \quad M_\lambda(T) = \varepsilon(\lambda) M^0_\lambda(T)\,. \tag{14.13}$$

As an indication, $\varepsilon \approx 0.03$ for polished aluminium and 0.95 for lampblack. The emittance of a real body can thus be evaluated with reference to the emittance of a black body at the same temperature.

Note. Thermal wavelength measurements are always a delicate matter because any object will reflect IR radiation. In addition, the emissivity is very variable. Specifically, a hotter source placed in the neighbourhood of the object under investigation may lead to an illumination that would vitiate the measurement result.

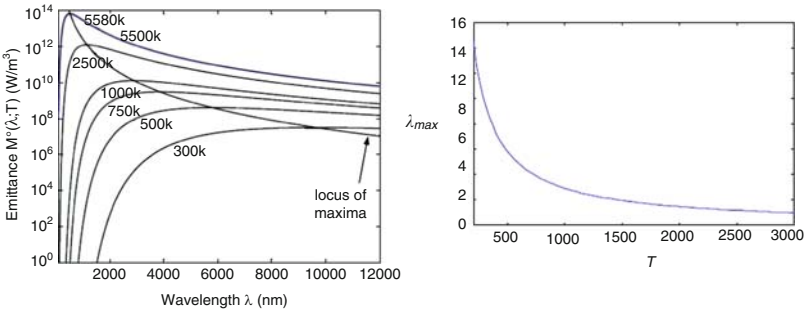

Fig. 14.5 Thermal radiation. (**a**) Monochromatic emittance $M_\lambda^0(T)$ from a black body as a function of temperature. (**b**) Wavelength $\lambda_{\max}$ of the emission maximum as a function of temperature T (K)

When the detector can make measurements over a broad (optical) band, it is useful to integrate the Planck function. The integral over the whole spectrum is called the Stefan distribution, quantifying the total emission of the black body:

$$M(T) = \int_0^\infty M_\lambda^0(T)\mathrm{d}\lambda = \sigma T^4\,, \quad \sigma = 5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}K^{-4}}\,. \tag{14.14}$$

When the detector band is restricted to some range λ_1 to λ_2, only a fraction of this energy is recovered, viz.,

$$M_{\lambda_1-\lambda_2}(T) = \int_{\lambda_1}^{\lambda_2} M_\lambda^0(T)\mathrm{d}\lambda = F_{\lambda_1 T-\lambda_2 T}\,\sigma T^4\,, \tag{14.15}$$

where $F_{\lambda_1 T-\lambda_2 T}$ represents the recovered fraction of the black body emission.

14.1.3 Short Wavelength Measurements. Photon Flux

Far-field measurement of proper (intrinsic) emission is particularly well suited to determining the temperature of high temperature surfaces. It has the advantage that it relies only on perfectly understood physics, viz., the Planck distribution. However, as indicated in the introduction, the resolution of optical measurements is naturally limited by the diffraction phenomenon, according to the Rayleigh criterion $\Delta x \approx \lambda$. It is thus interesting to use short wavelengths, in the UV to visible range, to carry out the measurement, even if the energy emitted in this spectral range is very small [18]. As we shall show, another advantage is that the temperature measurement is generally more accurate at short wavelengths.

For short wavelength measurements, where one generally has $\lambda T \ll 14\,000\ \mu\mathrm{m\,K}$, the Wien approximation (14.16) to the Planck function can be applied:

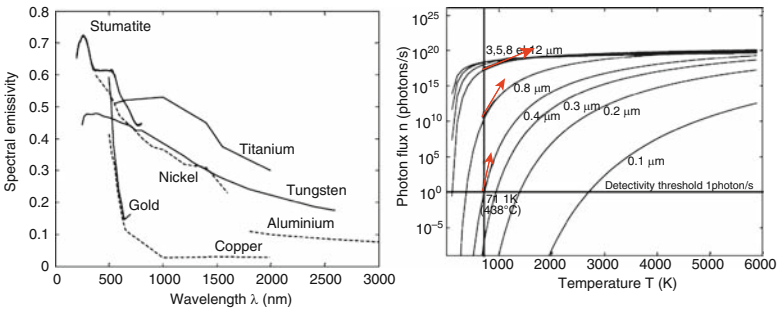

Fig. 14.6 Emissivity and photon flux at short wavelengths. (**a**) Emissivities of several materials as a function of wavlength. (**b**) Photon flux from a surface of equivalent dimensions to the wavelength of the detector

$$M_\lambda(T) = \varepsilon_\lambda C_1 \lambda^{-5} \mathrm{e}^{-C_2/\lambda T} \,. \tag{14.16}$$

One can then relate the true temperature T of an opaque surface and the luminance temperature T_L assuming the surface to be black, i.e., assuming $\varepsilon_\lambda = 1$:

$$\frac{1}{T} = \frac{1}{T_\mathrm{L}} + \frac{\lambda}{C_2} \ln \varepsilon_\lambda \,. \tag{14.17}$$

The difference between T and T_L thus decreases as λ is made smaller. Differentiating the Wien approximation and assuming that ε_λ is known with an error e_ε, it can then be shown that the error e_T in the temperature decreases as λ decreases and ε_λ increases:

$$\frac{e_\mathrm{T}}{T} = \frac{T_\mathrm{L} - T}{T} = -\lambda \frac{T}{C_2} \frac{e_\varepsilon}{\varepsilon_\lambda} \,. \tag{14.18}$$

Note that the relative error is proportional to the wavelength chosen. It therefore looks interesting to work in the UV–visible spectral range in order to make accurate temperature measurements. In switching from a measurement at 10 µm to one at 1 µm, the accuracy can in theory be improved by a factor of 10. Moreover, for many materials, the emissivity increases as the wavelength goes down [19], in particular for metal surfaces whose emissivity obeys Drude's law:

$$\varepsilon_\lambda = \alpha_\lambda = A \lambda^{-1/2} \,. \tag{14.19}$$

The error in the temperature will then be still smaller.

As the Planck distribution shows (see Fig. 14.5a), the energy emitted in the UV–visible band is very small and will be all the more so if the measurement surface is very small, or microscopic. At very short wavelengths, the flux is in fact so small that one no longer measures a continuous radiative flux, but rather a discrete photon flux. A photomultiplier must be used (see Sect. 14.2.1), together with a photon counting

technique [20]. The radiative flux expressed as an energy flux can be rewritten as a photon flux using the fact that the photon energy is $E = h\nu = hc/\lambda$. This new expression for the monochromatic flux $n_\lambda^{\mathrm{e}}(T)$ emitted by an area s of emissivity ε_λ can then be written

$$n_\lambda^{\mathrm{e}}(T) = \varepsilon_\lambda \frac{C_1 s}{hc} \lambda^{-4} \mathrm{e}^{-C_2/\lambda T} . \tag{14.20}$$

The temperature dependence of the photon flux for an emitting surface of diameter D ($s = \pi D^2/4$) of the same order of magnitude as the wavelength ($D \sim \lambda$) is given in Fig. 14.6b. The interest in making short wavelength measurements is now clear, because the variation of the signal about a given temperature becomes much greater. Indeed, it is this high sensitivity that explains the small error in the temperature measurement. However, there exists a threshold for photon detection below which it becomes impossible to use the results. For example, for a measurement at 0.4 μm on a surface of equivalent diameter, it will be impossible to measure temperatures below 711 K (438°C), because the theoretical maximum photon flux would then be only 1 photon/s.

14.1.4 Random Nature of the Photon Flux

Photon emission is a random quantum process [13]. Therefore the photon flux n cannot be measured exactly. However, the measurement remains accurate because the statistical laws obeyed by the photon flux are perfectly well known, viz., the Poisson distribution or normal distribution. Its value is in fact a random variable N related to the number of photons Y that the detector will count during a time Δt. If $\mu = \mathrm{E}(Y)$ and $n = \mathrm{E}(N)$ are the mathematical expectations of Y and N, respectively, then since $N = Y/\Delta t$, it follows that $\mu = n\Delta t$. The duration Δt is divided up into a large number q of very short time intervals δt during which the detector counts 1 photon (event $\{X = 1\}$, probability ϖ) or 0 photon (event $\{X = 0\}$, probability $1 - \varpi$). Thus X is a binary random variable, and Y the sum of all the independent random variables X, hence a binomially distributed random variable with integer expectation $\mu = q\varpi$. The probability of observing k photons ($0 \le k \le q$) thus has a binomial distribution:

$$\mathrm{Prob}\{Y = k\} = C_q^k \varpi^k (1 - \varpi)^{q-k} . \tag{14.21}$$

Since $\Delta t = q\delta t$, the probability ϖ is equal to $n\delta t$. The variance of Y is therefore

$$\mathrm{Var}(Y) = q\mathrm{Var}(X) = q\varpi(1 - \varpi) = \mu(1 - n\delta t) . \tag{14.22}$$

As can be seen from Fig. 14.7, if q is large ($q > 50$) and μ is less than 5, the binomial distribution tends to the Poisson distribution with parameter μ (expectation and variance), viz.,

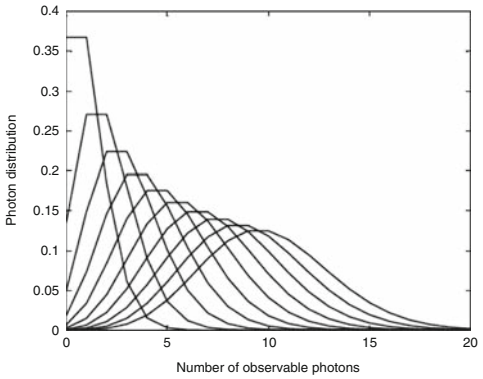

Fig. 14.7 Statistical distributions obeyed by the photon flux

$$\text{Prob}\{Y = k\} = \frac{\mathrm{e}^{-\mu}\mu^k}{k!} \qquad \text{(Poisson distribution)}\,, \tag{14.23}$$

while if μ is greater than 5, the binomial distribution tends to the normal distribution

$$\text{Prob}\{Y = k\} = \frac{1}{\sqrt{2\pi\mu}} \exp \frac{(k-\mu)^2}{2\mu} \qquad \text{(normal distribution)}\,. \tag{14.24}$$

However, whichever distribution is chosen, when the probability ϖ is small, one has $\text{Var}(Y) = \mu$. Consequently, over a duration Δt, a measurement of the expectation μ or the variance $\text{Var}(Y)$ of the signal yields the photon flux $n = \mu/\Delta t$.

So whatever the level of the photon flux and whatever statistical distribution governs it, the temperature can always be found either from the expectation ('average') n of the photon flux N, or from its variance ('standard deviation'), or better still, from both of them together and (14.20).

14.1.5 Multispectral Measurements

The radiative or photon flux $n_\lambda^{\mathrm{r}}(T)$ measured experimentally by the detector or camera (received flux n_λ^{r}) through a voltage or current measurement is proportional to the flux emitted by the surface:

$$n_\lambda^{\mathrm{r}}(T) = \mathrm{TF}_\lambda k n_\lambda^{\mathrm{e}}(T) = \varepsilon_\lambda \frac{C_1 s}{hc} \mathrm{TF}_\lambda k \lambda^{-4} \mathrm{e}^{-C_2/\lambda T}\,, \tag{14.25}$$

where TF_λ is the transfer function of the microscope, depending on the spectral transmittivities of the various elements making it up, viz., transmittivities of the objective and eyepiece of the microscope, and the quantum efficiency of the PMT, and k is a coefficient of proportionality that is independent of the wavelength and

takes into account purely geometrical features such as the numerical aperture, shape factors related to the throughput, emission area, and so on.

There are therefore two operations in a determination of the surface temperature. First, a calibration of the measurement setup to determine the transfer function TF_λ and the coefficient of proportionality k, and then a measurement of the local emissivity ε_λ of the surface. This second step is a delicate matter on the microscale. Moreover, ε_λ is a parameter that is likely to vary in space and time (due to oxidation, accumulation of dirt on the surface, etc.). It is thus worthwhile trying to get round this double difficulty by carrying out multispectral measurements, using a set of monochromatic filters that are close to one another in wavelength.

The idea here is to work with the ratio of the photon fluxes n_{λ_i} and n_{λ_j} at two wavelengths λ_1 and λ_2 [21]. In order to justify the assumption of constant emissivity $\varepsilon_{\lambda_i} \approx \varepsilon_{\lambda_j}$ and transfer function $\mathrm{TF}_{\lambda_i} \approx \mathrm{TF}_{\lambda_j}$, the two wavelengths must be chosen close to one another. Under these conditions, the temperature can be retrieved using the following relation:

$$T = \frac{C_2(1/\lambda_i - 1/\lambda_j)}{\ln\left[\dfrac{n_{\lambda_i}}{n_{\lambda_j}}\left(\dfrac{\lambda_j}{\lambda_i}\right)^{-4}\right]} \,. \tag{14.26}$$

The relative error e_T/T in the temperature is deduced by differentiating the photon flux ratio and treating the differential terms as errors:

$$\frac{e_\mathrm{T}}{T} = \frac{\dfrac{e_{\lambda_i}}{\lambda_i}\left|5 - \dfrac{C_2}{\lambda_i T} - \dfrac{\lambda_i}{\varepsilon}\dfrac{\mathrm{d}\varepsilon}{\mathrm{d}\lambda}\right| + \dfrac{e_{\lambda_j}}{\lambda_j}\left|5 - \dfrac{C_2}{\lambda_j T} - \dfrac{\lambda_j}{\varepsilon}\dfrac{\mathrm{d}\varepsilon}{\mathrm{d}\lambda}\right| + \dfrac{e_{n_{\lambda_i}}}{n_{\lambda_i}} + \dfrac{e_{n_{\lambda_j}}}{n_{\lambda_j}}}{\dfrac{C_2}{T}\left|\dfrac{1}{\lambda_i} - \dfrac{1}{\lambda_j}\right|} \,. \tag{14.27}$$

This relation shows that the error in the temperature increases as the wavelengths λ_i and λ_j come closer, because the denominator then tends to zero. To avoid amplifying the error, this term must therefore be bigger than 1.

The optimal choice is therefore a compromise between the assumption of constant emissivity and an error amplification less than 1. For example, for $\lambda_i \approx \lambda_j \approx 0.4\ \mu\mathrm{m}$ and $T = 1\,000\ \mathrm{K}$, one has $|\lambda_i - \lambda_j| > 10\ \mathrm{nm}$.

14.2 Measurement by Photomultiplier and UV to NIR Camera

In short wavelength thermography, we measure a fraction of the spectral distribution of the flux radiated by the object whose temperature is to be measured. Because we work at short wavelength, the detected flux is naturally very small, but this is compensated for by two key points:

- The sensitivity is exponential, which leads to excellent temperature resolution, provided that the signal-to-noise ratio (SNR) is acceptable. It follows that the effect of surface emissivity is much less critical for these measurements.
- The spatial resolution is naturally associated with the wavelength, and the significant gain here allows one to localise the measurement, which is useful for microsystems.

For narrow band detection, the output voltage of the photodetector can be written in the form

$$V(\lambda, T) = a \int_{\lambda}^{\lambda+\Delta\lambda} n_{\lambda}^{\mathrm{r}}(T)\mathrm{d}\lambda \ , \tag{14.28}$$

where a contains the quantum efficiency of the detector and the current–voltage conversion factor (often a simple resistance, in which case the factor is the value of the resistance).

Note that the relation is not the same as for total radiation pyrometers, which are based on the Stefan law. Given the significant variation of the number of emitted photons for different wavelengths and the narrow spectral band, the surface emissivity and optical transmission factor can often be treated as constant. With this hypothesis, provided it can be validated, the measurement can be greatly simplified.

14.2.1 Principle of Photomultipliers and Cameras

These two kinds of detector use the interaction between photons and electrons, and are called quantum photodetectors, because it is the light quantum that produces electrical charges. They have two fundamental theoretical properties:

- The photodetector will be able to measure the radiation if the wavelength λ is shorter than the threshold wavelength $\lambda_{\text{threshold}}$, which depends on the work function of the material.
- The efficiency of the photodetector is phenomenologically proportional to the wavelength if $\lambda < \lambda_{\text{threshold}}$.

Generally speaking, when choosing a photodetector, several parameters need to be considered, namely, the working spectral band, the sensitivity, the dynamic range (hence the noise level), and the temporal response (integration time, response time). In general, the sensitivity depends on λ and is therefore associated with the spectral band. The spectral sensitivity is defined as

$$S_i(\lambda) = \left.\frac{\partial i}{\partial \Phi}\right|_{\lambda} \ , \tag{14.29}$$

where i is the current and Φ the measured flux (in the sense of the optical power).

The noise level of the photodetector is given by the noise equivalent power (NEP). It is defined for a noise band of 1 Hz. At this level, the quality of the photodetector is indicated by the specific detectivity D^*, defined by

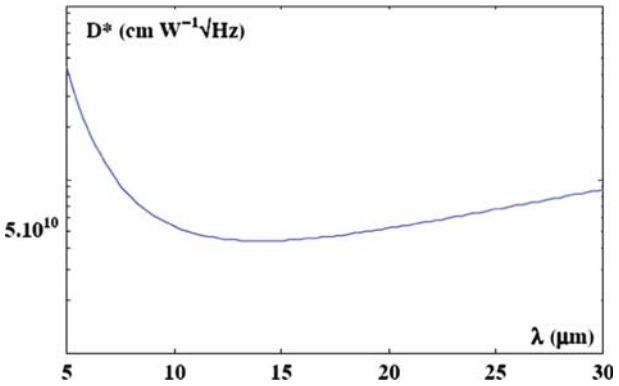

Fig. 14.8 Theoretical limit of D^* as a function of the wavelength for a quantum photodetector

$$D^* = \frac{\sqrt{A}}{\text{NEP}},$$

where A is the surface area of the detector. D^*, which is usually given in units $\text{W}^{-1}\text{cm}\,\text{Hz}^{1/2}$, fixes the experimental measurement limit in optimal conditions. Theoretically, it has a minimum for a wavelength of 14 μm, as can be seen from the theoretical curve in Fig. 14.8.

In the present chapter, we shall limit the discussion to two types of quantum photodetector, viz., the photomultiplier and the CCD camera.

Photomultipliers

A photomultiplier (PMT) is an association of a photoemitting cell and an amplifier [22, 23]. A flow of electrons in vacuum is exploited to amplify directly using secondary emission electrodes called dynodes. The dynodes are biased by a high voltage supply, so that the potential difference between successive dynodes is of the order of 100 V (except in integrated PMTs).

Figure 14.9 illustrates how amplification is brought about inside the photomultiplier:

- Amplification in one step $n_t\delta$, where η_t is the transfer efficiency and δ the secondary emission coefficient (number of secondary electrons emitted for one incident electron).
- Total gain $G_i = \eta_c$, where η_c is the collection efficiency in the first dynode, and n is the number of dynodes.

Note that δ depends on V (often roughly proportional to V). The photomultiplier gain can thus be adjusted by modifying the supply voltage.

Figure 14.10 shows how a photomultiplier can be biased using a voltage divider bridge. In this setup, the Zener diode (ZD) fixes the operating point of the

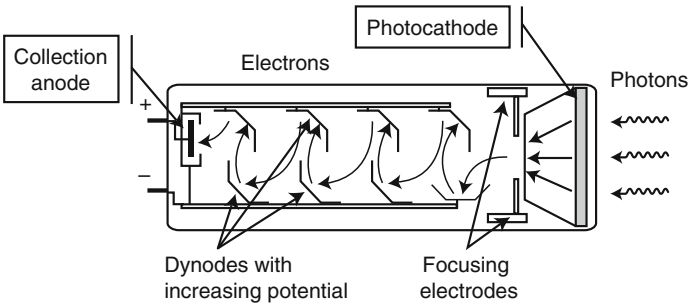

Fig. 14.9 Amplification in a photomultiplier

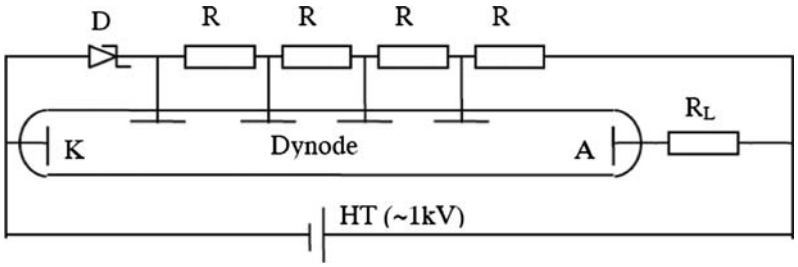

Fig. 14.10 Practical arrangement for biasing a photomultiplier

photocathode and hence also the efficiency η_c and sensitivity S_i. If necessary, capacitors are connected in parallel with the resistances of the last dynodes to bring about decoupling (in the case of pulsed fluxes or strong transient currents).

The main features of photomultipliers are as follows:

- High gain amplifier (typically 50–140 dB) with no specific thermal noise, but a source of $1/f$ noise.
- Wide pass band (> 100 MHz) and high specific detectivity ($\sim 10^{16}\,\mathrm{W^{-1}cm\,Hz^{1/2}}$ for 0.4 μm and $-145°$C).

Charge Coupled Devices (CCD)

Charge transfer circuits arose from work carried out independently by research at Philips, which led to the bucket brigade devices (BBD) in 1969, and research at the Bell Laboratories, which led to the charge coupled devices (CCD) in 1970. The latter have been considerably developed since then. CCDs face serious competition from CMOS technology in everyday applications. However, the dynamic range of the CCD remains a major strong point in measurement applications.

The basic idea of the CCD is depicted in Fig. 14.11 [24]. The structure is of MOS type and the substrate type *N*. Holes (positive charges) are generated by the action of photons between the struts of the metal grid. If one of the electrodes is held at a negative potential, then a potential well results under the surface, just waiting to be

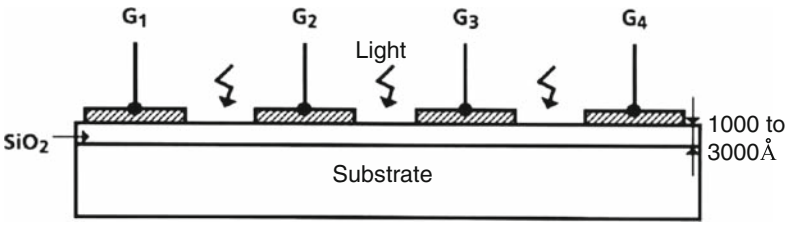

Fig. 14.11 Structure of a CCD

'filled' with holes. The charges are stored there (as in a capacitor), and their numbers will not be affected over some short time scale (shorter than the time required for the inversion layer to form). Charge transfer is achieved by emptying the capacitors into their neighbours. This transfer is controlled by logical signals generated by a clock and applied to each of the electrode interconnects. So for example, a charge produced at the first electrode moves forward by small jumps toward the output diode, where it is measured. This propagation is synchronised with the clock signals. Between the CCD input and output, the signal is delayed by a time equal to the product of the clock period and the number of steps in the CCD.

The main problem with a CCD transmission line is charge loss due to transfer over a large number of elements. To quantify this effect, we define the transfer inefficiency:

$$\varepsilon = \frac{Q_\mathrm{r}}{Q_0}\,, \tag{14.30}$$

where Q_0 is the charge to be transferred and Q_r the charge not transferred.

CCDs are usually made from silicon, which is photosensitive in a spectral band from the visible to the near infrared (NIR). The sensor is placed in the focal plane of an optical device which produces an image of the scene under observation. After scanning the array, a video signal is reconstituted electronically, then often converted before sending to a microcomputer. The main features of a CCD camera are the type of sensor, the size of the array, the spectral sensitivity, the dynamic range of the video signal (especially after conversion), uniformity of the array (a key feature in metrology), and a spatial resolution depending on the optics.

14.2.2 Experimental Setup for the UV Thermal Microscope

Experimental Arrangement

With the aim of making microscale temperature measurements by photon counting (see Sects. 14.1.3 and 14.1.4), the ultraviolet microscope (see Fig. 14.12) developed by LEMTA (France) [25, 26] comprises a CaF_2 UV microscope objective with a long working distance (15 mm) ($\times$50, numerical aperture NA = 0.42), a BK7

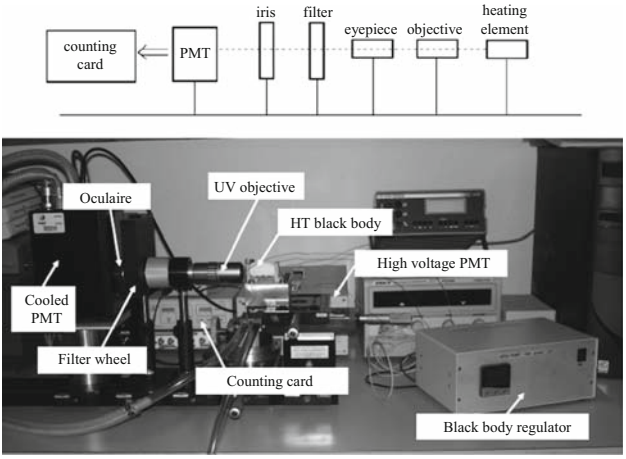

Fig. 14.12 Photon-counting multispectral UV microscope for temperature measurements

($\times 10$) eyepiece transparent from 350 nm, and a diaphragm placed between the eyepiece and microscope objective to control the spatial resolution of the target region. The whole setup is isolated from the surroundings by a reflecting copper tube. Focusing is achieved using an XYZ microdisplacement system. The Multi-Alkali photomultiplier is cooled by the Peltier effect (noise < 1 photon/s). It has a spectral sensitivity in the range 0.185–0.710 μm. Its maximal efficiency is reached in a band between 250 and 450 nm. The photons collected by the PMT are transformed into photoelectrons by the photocathode, and a series of dynodes at increasing potentials serves to amplify the signal. These photoelectrons are then counted using a discrimination and counting card.

An iris diaphragm and three monochromatic filters ($\overline{\lambda} = 380$, 390, and 400 nm, with $\delta\lambda = 10$ nm) are placed in front of the PMT, mounted on a filter wheel. The quantum efficiency η of the PMT, and the transmittivities τ_{obj} and τ_{eye} of the objective and eyepiece, respectively, are functions of the wavelength λ. The product of these quantities and the transmittivity τ_λ of the monochromatic filter is denoted by TF_λ as in (14.25). Figure 14.13a shows the transfer function TF_λ of the microscope and Fig. 14.13b the transmittivities τ_λ of the monochromatic filters used for the multispectral measurement. Note that the microscope transfer function is maximal and relatively constant over a spectral range corresponding to the 3 filters, viz., 370–410 nm.

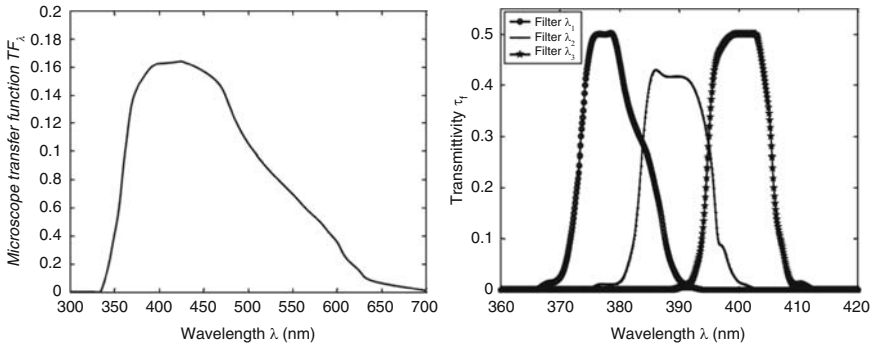

Fig. 14.13 (**a**) Transfer function TF_λ of the microscope and (**b**) spectral transmittivities τ_λ of the three filters

Application to the Multispectral Method

Choice of Filters

As discussed in Sect. 14.1.5, the optimal choice for the filters results from a compromise between an amplification of the error in the temperature and the assumption of constant transfer function and emissivity. When nothing is known about the emissivity, it is difficult to provide a quantitative answer to this problem. The three filters of this microscope were chosen to coincide with the maximum of the transfer function. We also imposed a constraint on the signal ratios (at least a factor of 2 for two consecutive filters). The wavelength ratio λ_i/λ_j is around 0.975, which corresponds to the wavelengths 380, 390, and 400 nm chosen for the filters. It is interesting to note that the wavelengths are only 10 nm apart, something made possible by the high sensitivity of the signal available in the near-UV to visible range. Note also that this is altogether compatible with the result concerning the amplification of the error in the temperature discussed in Sect. 14.1.5.

Non-Ideal Multispectral Method

Experimentally, the microscope transfer function cannot be considered as strictly constant. Moreover, the filters are not truly monochromatic, but can be characterised by a central wavelength $\overline{\lambda}$, a pass band $2\delta\lambda$, and a maximal transmittivity τ (see Fig. 14.13b). These features must be taken into account in the expression for the measured photon flux, which must include not only the transfer function TF_λ of the microscope and geometrical factors k, but also the pass band $\overline{\lambda} \pm \delta$ of the filters and their transmittivity τ_λ, giving

$$\overline{n}^{\mathrm{r}}_{\overline{\lambda}}(T) = \frac{C_1 s}{hc} \int_{\overline{\lambda}-\delta\lambda}^{\overline{\lambda}+\delta\lambda} \varepsilon_\lambda \mathrm{TF}_\lambda k \tau_\lambda \lambda^{-4} \mathrm{e}^{-C_2/\lambda T} \mathrm{d}\lambda \ . \tag{14.31}$$

A good (second order) approximation to this is given by

$$\overline{n}^{\mathrm{r}}_{\overline{\lambda}}(T) = \varepsilon_{\overline{\lambda}} \mathrm{TF}_{\overline{\lambda}} k \tau_{\overline{\lambda}} \frac{C_1 s}{hc} \int_{\overline{\lambda}-\delta\lambda}^{\overline{\lambda}+\delta\lambda} \mathrm{d}\lambda = 2\delta\lambda \varepsilon_{\overline{\lambda}} \mathrm{TF}_{\overline{\lambda}} k \tau_{\overline{\lambda}} \frac{C_1 s}{hc} \overline{\lambda}^{-4} \mathrm{e}^{-C_2/\overline{\lambda}T} \ . \tag{14.32}$$

The ratio of the photon fluxes given by (14.32) is obtained for two filters with average wavelengths λ_i and λ_j. It now depends only on TF_λ (the transfer function of the microscope) and τ_λ (the transmittivity of the filter), which are both known, whence the temperature T can be recovered using

$$\frac{\overline{n}^{\mathrm{r}}_{\lambda_i}(T)}{\overline{n}^{\mathrm{r}}_{\lambda_j}(T)} = \frac{\mathrm{TF}_{\lambda_i}}{\mathrm{TF}_{\lambda_j}} \frac{\tau_{\lambda_i}}{\tau_{\lambda_j}} \left(\frac{\lambda_i}{\lambda_j}\right)^{-4} \frac{\delta\lambda_i}{\delta\lambda_j} \exp\left[-\frac{C_2}{T}\left(\frac{1}{\lambda_i}-\frac{1}{\lambda_j}\right)\right] \frac{\varepsilon_{\lambda_i}}{\varepsilon_{\lambda_j}} \ , \tag{14.33}$$

where the last factor $\varepsilon_{\lambda_i}/\varepsilon_{\lambda_j} \approx 1$, and

$$T = \frac{C_2\left(\dfrac{1}{\lambda_j}-\dfrac{1}{\lambda_i}\right)}{\ln\left[\dfrac{\overline{n}^{\mathrm{r}}_{\lambda_i}}{\overline{n}^{\mathrm{r}}_{\lambda_j}} \dfrac{\mathrm{TF}_{\lambda_j}}{\mathrm{TF}_{\lambda_i}} \dfrac{\tau_{\lambda_j}}{\tau_{\lambda_i}} \dfrac{\delta\lambda_j}{\delta\lambda_i} \left(\dfrac{\lambda_j}{\lambda_i}\right)^{-4}\right]} \ . \tag{14.34}$$

The different parameters $\overline{n}$, TF, τ, and $\delta\lambda$ in the denominator must be known for the chosen wavelengths λ_i and λ_j, and therefore require optical characterisation of the filters and determination of the microscope transfer function. Many precautions are taken to optimise the measurement: the filters are chosen close in wavelength $\overline{\lambda}$ to be able to treat the emissivity ε_λ as constant, with almost the same nominal characteristics $(\tau, \delta\lambda)$ and with average wavelengths $\overline{\lambda}$ centered on a spectral range where the microscope transfer function TF_λ varies only slightly. It is thus easy to see why poor knowledge or measured values of these quantities might lead to non-negligible errors in the temperature.

Three-Wavelength Multispectral Method

The measurement can be improved by taking into account the variation of the microscope transfer function, the emissivity, and the spectra of the filters by means of a correction function $f(\lambda)$. Since the wavelengths of the three filters (i, j, k) are close together, this function can be replaced by an order 2 expansion about $\lambda_\mathrm{m} = (\lambda_i + \lambda_j + \lambda_k)/3$, viz.,

$$f(\lambda) = f(\lambda_\mathrm{m}) + (\lambda - \lambda_\mathrm{m}) f'(\lambda_\mathrm{m}) + \frac{1}{2}(\lambda - \lambda_\mathrm{m})^2 f''(\lambda_\mathrm{m}) + O\big((\lambda - \lambda_\mathrm{m})\big)^2 \ . \tag{14.35}$$

For the filters i and j, (14.34) becomes

$$T = \frac{C_2\left(\frac{1}{\lambda_j} - \frac{1}{\lambda_i}\right)}{\ln\left[\frac{\overline{n}^{\mathrm{r}}_{\lambda_i}}{\overline{n}^{\mathrm{r}}_{\lambda_j}} \frac{1+(\lambda_j-\lambda_\mathrm{m})X+(\lambda_j-\lambda_\mathrm{m})^2 Y}{1+(\lambda_i-\lambda_\mathrm{m})X+(\lambda_i-\lambda_\mathrm{m})^2 Y}\left(\frac{\lambda_j}{\lambda_i}\right)^{-4}\right]}\,, \tag{14.36}$$

where

$$X = \frac{f'(\lambda_\mathrm{m})}{f(\lambda_\mathrm{m})}\,, \qquad Y = \frac{f''(\lambda_\mathrm{m})}{2f(\lambda_\mathrm{m})}\,.$$

This yields the temperature T from the photon flux measurements $\overline{n}^{\mathrm{r}}_{\lambda_i}$ and $\overline{n}^{\mathrm{r}}_{\lambda_j}$ through filters i and j, but X and Y are also unknown. A new filter k at a new wavelength λ_k must therefore be introduced. Then from the 3 photon flux ratios $\overline{n}^{\mathrm{r}}_{\lambda_i}/\overline{n}^{\mathrm{r}}_{\lambda_j}$, $\overline{n}^{\mathrm{r}}_{\lambda_i}/\overline{n}^{\mathrm{r}}_{\lambda_k}$, and $\overline{n}^{\mathrm{r}}_{\lambda_j}/\overline{n}^{\mathrm{r}}_{\lambda_k}$, and three applications of (14.36), a non-linear system of three equations in three unknowns is obtained, whence the values of X, Y, and T can be determined.

Experimental Results

Figure 14.14 shows the validation results obtained with this microscope using two calibration devices (see [26] for more details concerning these devices). The temperature measurements were made on a microscopic region. Figure 14.14a shows the dependence of the photon flux measured by the PMT (0.185–0.710 μm) on the temperature of an Alumel wire heated by the Joule effect. The temperature of the wire is determined by measuring its resistance. The integration time was 50 ms. We then fitted the Planck law which described this temperature dependence of the photon flux. The experimental results agree perfectly with theory. Figure 14.14b shows the photon fluxes measured on a double cavity ceramic (Stumatite) black body developed in the lab. The temperature of the black body is controlled by means of two thermocouples placed on either side of the cavity. In our experiment, the temperature of the thermocouples was about 1 000°C. Experimentally, the objective was focused on a fictional surface located at the entrance of the second hole, which has a diameter of micron order. The photon counting card counts the number Y of photoelectrons detected over a time Δt. Since the count is random, one can calculate the average number of photons $\mu = \mathrm{E}(Y)$ over the time Δt, then the photon flux $n = \mu/\Delta t$, either from the expectation N or from the variance V_a of the photon flux (standard deviation $\sigma = \sqrt{V_\mathrm{a}}$), and this for each of the filters (see Sect. 14.1.4).

Table 14.1 shows the photon counts $\overline{n}^{\mathrm{r}}_{\lambda}(T)$ for the three filters labelled 1, 2, and 3. The temperatures T are obtained using (14.26) for T_{NTF} (without transfer function TF), then (14.34) to obtain T_{WTF} (with transfer function), and finally (14.36). This shows that it is important to have information about the microscope transfer function in order to make accurate temperature measurements. The three-wavelength method is able to correct for any variations in these parameters at different wavelengths. We

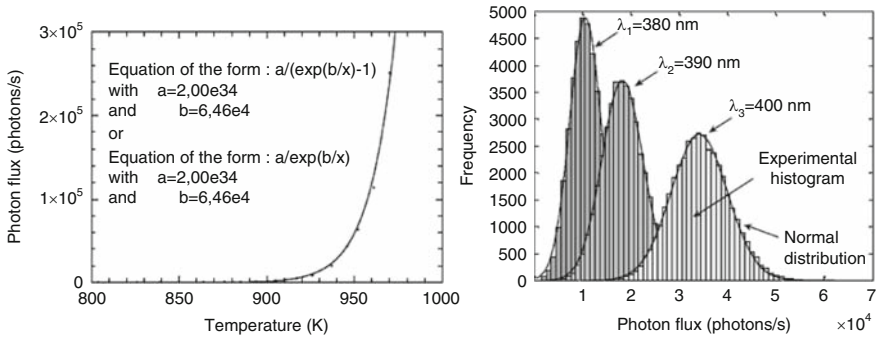

Fig. 14.14 Experimental results obtained with the UV microscope. (**a**) Measured photon flux as a function of the temperature of the Alumel wire. (**b**) Photon flux measured by the 3 monochromatic filters (double cavity black body)

Table 14.1 Temperature results with the different techniques

$\overline{n}^{\mathrm{r}}_{\lambda}(T)$ [photons/s]	Ratio	T_{NTF} [°C]	T_{WTF} [°C]	T [°C]	X	Y
10 536	i/j	1 298	1 088			
18 221	i/k	1 184	1 075	1 075	-1.4598×10^{-4}	1.0863×10^{14}
34 132	k/j	1 082	1 063			

find

$$\sup_{n=i,j,k}\left|\frac{(\overline{\lambda}_n-\lambda_{\mathrm{m}})f'(\lambda_{\mathrm{m}})}{f(\lambda_{\mathrm{m}})}\right|=3.78\times10^{-4}\,,\quad \sup_{n=i,j,k}\left|\frac{(\overline{\lambda}_n-\lambda_{\mathrm{m}})^2f''(\lambda_{\mathrm{m}})}{2f(\lambda_{\mathrm{m}})}\right|=0.0125\,.$$

It turns out that $f'(\lambda_{\mathrm{m}})$ is close to zero. This implies that we are indeed close to the maximum of the transfer function. The constant emissivity hypothesis thus seems to be valid, since this maximum has not shifted.

14.2.3 Experimental Setup for the Silicon CCD Camera

Experimental Arrangement

The radiometric method using a CCD camera developed at FEMTO-ST (France) [27] was mainly devoted to thermal studies of microstructures, because a spatial resolution close to the optical resolution (around 500 nm) can be obtained by working

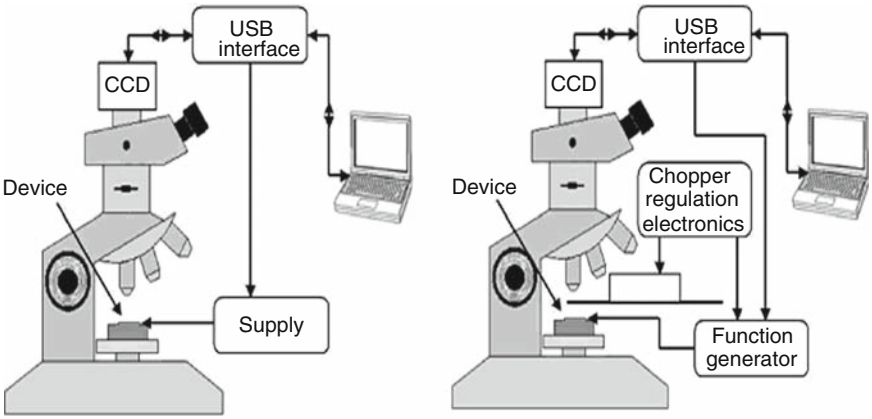

Fig. 14.15 Experimental arrangements for thermography by CCD camera. (**a**) Static detection. (**b**) Dynamic detection

in the visible–near infrared wavelength range (roughly 0.8–1 μm). As mentioned above, it also allows highly sensitive detection of small variations in thermal emission corresponding to an average temperature of the order of 300°C, or less in some cases. This multispectral technique is particularly well suited to static or very low frequency temperature measurements. Two arrangements have been used depending on the kind of detection:

- The static measurement technique (see Fig. 14.15a) is particularly simple since it uses an optical microscope (LEICA) in association with a digital camera. (We used different models: DALSA fast but not cooled, Hamamatsu and Kodak cooled. The latter two devices have very low noise and high dynamic range.) The interface electronics was specifically designed to allow the high dynamic range needed for the measurement (ideally 100 dB) and a USB link with a microcomputer.
- The experimental arrangement set up for dynamic detection is illustrated in Fig. 14.15b. A mechanical shutter device (chopper) consisting of a rotating disk was included in the observation line of the microscope. The rotating disk is equipped with an optical sensor for regulating the speed and synchronising the function generator supplying the heating device under investigation. This mechanical chopping device limits the frequency to around 10 kHz.

Image Processing, Representation, and Calibration

Image Processing

For radiometric measurements, image processing is unnecessary when the temperature remains constant. However, under dynamic conditions, such processing is unavoidable in order to extract the amplitude and phase of the temperature at each point, since this provides access to the thermal properties of the material. The amplitude and phase can be calculated from a digital lock-in detection, the simplest involving the measurement of 4 points per period (quadrature between two measurements). Suppose the temperature varies sinusoidally, and let d_i be the contents of the pixels of the four images taken successively. The amplitude A_T and phase φ are given simply by

$$A_\mathrm{T} = \frac{1}{2}\sqrt{(d_1 - d_3)^2 + (d_2 - d_4)^2}\,, \qquad \varphi = \arctan\frac{d_1 - d_3}{d_2 - d_4}\,. \tag{14.37}$$

Processing thus consists in going from 4 images to 2 images using these expressions. It is then easy to reconstruct the temperature image at any time. The image can be represented in false colours by appealing to the calibration curve, viewed as a look-up table (LUT) for the user.

The signal-to-noise ratio (SNR) is one of the fundamental pieces of data for this measurement system. Indeed, this is what fixes the ultimate measurement limit. Specifically, the CCD camera integrates, and in the case of a static measurement, it can be used basically with long integration times, since the SNR has the following simple form:

$$\mathrm{SNR} = \frac{St}{\sqrt{\sigma_\mathrm{p}^2 + \sigma_\mathrm{D}^2 + \sigma_\mathrm{r}^2}}\,, \tag{14.38}$$

where S represents the signal power (number of electrons generated each second), t is the integration time, σ_p is the photon noise, which can be neglected for low detection levels, σ_D is the dark noise which depends on the temperature (this noise, given by Meyer's relation, can be neglected if the camera is cooled), and σ_r is the read noise, which is generally the most critical for our application (noise depending on the emptying frequency).

It is useful to exploit the physical properties of the noise in order to improve the signal-to-noise ratio. A second image processing sequence was therefore considered extremely worthwhile, given the potential for improvement that can be achieved by binning. (This technique works out a spatial average by summing the charges in a region of the camera. This sum is carried out directly before transfer to the microcomputer.) Indeed, by binning, the signal-to-noise ratio is increased in proportion to the number of elements averaged, because the sum is carried out before reading. This can increase the number of photons generated, corresponding to S in (14.38), by increasing the equivalent area without changing the microscope objective. The potential weak point in this technique is a degradation of the spatial resolution, although it can be restored under two conditions: the object must be viewed in motion

(it is the motion that makes superresolution possible), and it is better if the motion is uniform over the whole surface (so-called rigid motion). The details of this kind of image processing, leading to superresolution, go beyond the scope of this overview, and the reader is referred to the literature [27–29].

Calibrating the Microthermograph

One of the difficulties raised by microradiometry has been calibration. We have developed two methods for achieving this [30]. The first uses a heating track equipped with a heat sensor. (A simple approach might have been to use a deposited platinum track whose electrical resistance would have served simultaneously as the sensor, but this method is inaccurate because of possible temperature gradients.) We chose to use a microthermocouple as target in order to remove the twofold uncertainty concerning the emissivity and the temperature. Specifically, the high resolution of the microscope led to the design and fabrication of a customised microdevice with a surface temperature that can be very precisely regulated by electrical heating. The calibration was carried out in two stages, because the heat sink resulting from the presence of the thermocouple necessarily induces a local temperature variation. To begin with, the optical response of the camera was measured with a large-area sample (silicon substrate coated with a chromium film), whose emissivity had been measured previously. The sample was then heated by a flat resistance serving as support, assuming that the emissivity remained constant.

The second calibration method, more delicate to implement, is based on the fabrication of a microscale black body. Indeed, a conventional black body would be unsuitable due to its large dimensions. A microdevice was made (aperture 1 mm) in order to calibrate the thermographic microscope. The reference surface was the flat and blackened surface of a thermocouple junction heated by a coil with a current through it [28].

Measurement Limits and Optimisation

The experimental measurement limit naturally depends on the total noise level of the CCD sensor. This measurement limit corresponds to the temperature, which corresponds in turn to SNR = 1 for one pixel. Of course, the minimal measurable temperature depends on the emissivity and integration time. Here we shall only discuss the effect of the first parameter, which is the most critical, because it is independent of the operator. The signal-to-noise ratio for a pixel is given simply by (the photon noise is proportional in power to the signal):

$$\mathrm{SNR} = \frac{St}{\sqrt{St + \sigma_\mathrm{D}^2 + \sigma_\mathrm{r}^2}} \,. \tag{14.39}$$

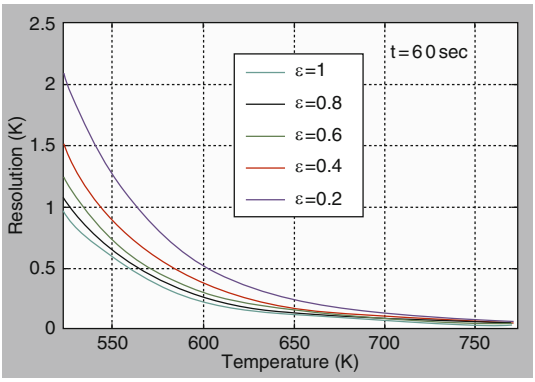

Fig. 14.16 Temperature dependence of the ultimate resolution, for different values of ε

Table 14.2 Lowest detectable temperatures for different values of t and ε

t [s]	$\varepsilon = 0.2$	$\varepsilon = 0.6$	$\varepsilon = 1$
10	502	483	474
20	493	474	466
60	481	463	455
100	476	458	451

The read noise is constant. For its part, the dark noise can be neglected to a first approximation when the camera is cooled. Regarding the temperature resolution, the calculated results shown in Fig. 14.16 correspond to a camera equipped with the Hamamatsu type S7031 sensor. The integration time was chosen as 60 s for the calculation. Note that the detection limit varies only slightly around this already long value, as can be seen from Table 14.2, which gives the detection thresholds for different values of the emissivity and integration time.

Applications to MEMS

The sample designed to validate this method is shown in Fig. 14.17. It consists of a grating of chromium tracks of width 12 μm, with a grating interval of 17 μm. The chromium was deposited on a passivated silicon substrate at the FEMTO-ST technology center (France). Chromium has emissivity 0.7, assumed constant for the measurement.

The grating was heated by the Joule effect at a constant current. The device was tested for different power values. Figure 14.18 (left) shows the temperature distribution at the surface for a power of 10.7 W. The image contrast is clearly due to the thermal emission of the conductors, whose central temperature is 660 K. As expected, the maximal temperature lies in the central part of each track (heat conduction

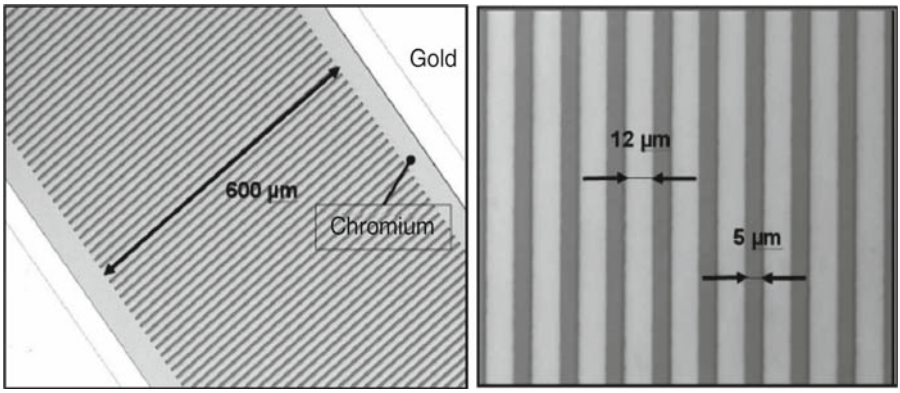

Fig. 14.17 Optical images of the system of conducting chromium tracks

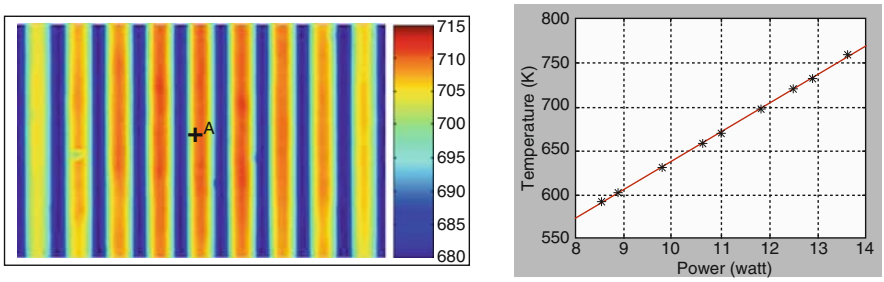

Fig. 14.18 *Left*: Temperatures of the tracks as measured by radiometry (scale in K). *Right*: Evaluation of the temperature at the point A

by the substrate and convection in the air). The temperature of the regions between the tracks is approximate because the calculation was carried out using the chromium emissivity everywhere. Figure 14.18 (right) shows the linear dependence of the temperature calculated as a function of the supply power, in agreement with the linear model of the device. The spatial resolution was around 400 nm and the temperature resolution 0.1 K depending on T.

14.3 Conclusion

We have been able to demonstrate the advantages of working with short wavelengths to improve the spatial resolution and accuracy of temperature measurements by intrinsic emission. The far-field radiometry methods discussed in this chapter are perfectly complementary to near-field optical methods, especially with regard to high temperature measurements. Moreover, these are largely non-intrusive methods, providing a rather easy way of deducing the temperature, since the photon flux due to intrinsic emission is not much affected by surface topology.

Regarding the UV microscope, we have shown that it is useful to implement a multispectral method to obviate the need for calibration or measurement of the local emissivity of the surface, two operations that are difficult to achieve in practice, especially for high temperature surfaces where the emissivity can easily vary both in time and from one point to another on the surface. Furthermore, there are few temporally reliable black bodies at high temperatures for characterising this kind of detector. Several difficulties came to light. In particular, it was shown that the multispectral measurement was not totally independent of the microscope transfer function, or indeed of the optical properties of the monochromatic filters. One solution for reducing the measurement error is to carry out a three-band multispectral measurement. The main difficulty that remains here comes from the characteristics of these filters, which are not strictly identical. One prospect for development that is currently under investigation at LEMTA would be to replace these three filters by a multichannel analyser or a diffraction grating. In the latter case, the same optics is used for the three measurements, which solves the problem of the non-ideally monochromatic filters. Finally, to lower the detectivity threshold of passive radiometric methods, which can be prohibitive for some applications, one solution would be to investigate stimulated photothermal methods, such as laser-induced fluorescence [31].

For its part, camera thermography has developed very rapidly over the past few years with the commercial availability of high quality and low cost cameras, thanks to the generalised integration of components. For room temperature applications, the cameras used are naturally infrared (few photons in the near infrared and visible), but it is clear that above 300°C, the CCD camera becomes extremely effective, and it should witness a general expansion due to two key benefits: very high sensitivity, by virtue of Planck's law for short wavelengths, and high spatial resolution, typically 500 nm according to the Rayleigh criterion. This technique is well suited to small objects or highly localised measurements, and it should be rapidly extended to industrial applications. However, as with the other radiometric techniques, calibration remains a stumbling block. Indeed, the standard black bodies available are not well suited to local measurements, being too bulky, and new microscale standards will need to be developed to open the way to small scale thermal metrology. In addition, the emissivity problem, also encountered in the UV measurements, still needs careful attention for measurement applications, even though it is less critical for this wavelength range. One solution would obviously be to carry out a spectroscopic analysis of the emission surface at different temperatures, but the extra

cost involved would limit such measurements to specific objects involving perfectly identified materials and surface states.

A lot of work remains to be done in the field of short wavelength radiometric temperature measurements, before we will be able to develop an industrial measurement tool. But such a tool would be extremely beneficial in today's production processes.

References

1. Trannoy, N.: Des premiers microscopes à sonde locale au S.Th.M.(II): Historique de la microscopie thermique à sonde locale, Première école d'Hiver Micro et Nanothermique, Aussois (7–12 March 2004)
2. Majumdar, A.: Scanning thermal microscopy, Annu. Rev. Mater. Sci. **29**, 505–585 (1999)
3. Planck, M.: Über eine Verbesserung der Wienschen Spektralgleichung, Verhandl. Dtsch. phys. Ges. **2**, 202 (1900)
4. Gaussorgues, G.: *La thermographie infrarouge*, Technique et Documentation, Lavoisier (1984)
5. Tessier, G., Holé, S., and Fournier, D.: Quantitative thermal imaging by synchronous thermoreflectance with optimized illumination wavelengths, Applied Physics Letters **78**, 2267–2269 (2001)
6. Dilhaire, S., Grauby, S., and Claeys, W.: Calibration procedure for temperature measurements by thermoreflectance under high magnification conditions, Applied Physics Letters **84**, 822–824 (2004)
7. Jorez, S., Laconte, J., Cornet, A., and Raskin, J.P.: Low-cost optical instrumentation for thermal characterization of MEMS, Measurement Science and Technology **16**, 1833–1840 (2005)
8. Aizawa H., et al.: Fluorescence thermometer based on the photoluminescence intensity ratio in Tb doped phosphor materials, Sensors and Actuators A **126**, 78–82 (2006)
9. Bur, A.J., Roth, S.: Fluorescence temperature measurements: Methodology for applications to process monitoring, Polymer Engineering and Science **44**, 898–908 (2004)
10. Aigouy, L., De Wilde, Y., Mortier, M., Giérak, J., and Bourhis, E.: Fabrication and characterization of fluorescent rare-earth-doped glass-particle-based tips for near-field optical imaging applications, Applied Optics **43**, 3829–3837 (2004)
11. Taguchi, Y., Horiguchi, Y., Kobayshi, M., Saiki, T., and Nagasaka, Y.: Development of nanoscale thermal properties measurement technique by using near-field optics, JSME International Journal B **47**, 483–489 (2004)
12. De Wilde, Y., Formanek, F., Carminati, R., Gralak, B., Lemoine, P., Mulet, J.P., Joulain, K., Chen, Y., Greffet, J.-J.: Thermal radiation scanning tunnelling microscopy, Nature **444**, 740–743 (2006)
13. Rémy, B., Degiovanni, A., and Maillet, D.: Métrologie thermique dans l'ultraviolet. Application à la microthermique, RS série I2M, Hermes-Lavoisier, ISBN 2-7462-1316-8, Vol. 5, no. 1–2, 177–199 (2005)
14. Asch, G.: *Les capteurs en instrumentation industrielle*, 5th edn., Dunod, Paris (1998)
15. Broussaud, G.: *Optoélectronique*, Masson, Paris (1974)
16. Chaimowicz, J.C.: *Introduction à l'optoélectronique*, Dunod, Paris (1992)
17. Yariv, A.: *Optical Electronics*, 4th edn., Saunders College Publishing (1991)
18. Hervé, P., Sicard, J., and Rakotoarisoa, M.: Pyrométrie dans l'ultraviolet, Actes du colloque de thermique S.F.T., Belfort, 139–142 (1991)
19. Touloukian, Y.S.: *Thermal Radiative Properties*, Plenum (1970)

20. Hervé, P., Pinat, V.: Thermographie par comptage de photons dans l'ultraviolet, Deuxième colloque francophone: Méthodes et techniques optiques pour industrie, Vol. 1, S.F.O., Trégastel, 31–38 (2001)
21. Gardner, J.L., Jones, T.P., Davies, M.R.: A six wavelength radiation pyrometer, High Temperatures – High Pressures **13**, 459–466 (1981)
22. Hamamatsu Photonics K.K.: Photomultiplier tube: Principle to application, Manufacturer's documentation (1994)
23. Philips Photonics: Photomultiplier tubes, Manufacturer's documentation (1993)
24. Sequin, C.H., Thomssett, M.F.: *Charge Transfer Devices*, Academic Press, London (1975)
25. Pierre, T., Rémy, B., Degiovanni, A.: Micro-scale temperature measurement by multi-spectral and statistic method in the ultraviolet–visible wavelengths, J. Appl. Phys. **103** (1), 1–10 (2008)
26. Pierre, T.: Mesure de la température à l'échelle microscopique par voie optique dans la gamme ultraviolet-visible, PhD Thesis, Institut National Polytechnique de Lorraine, Nancy (2007)
27. Teyssieux, D.: Microscopie thermique et thermoélastique par caméra: Application à des microdispositifs, PhD Thesis, University of Franche-Comté, Besançon (2007)
28. Borman, S.: Topics in multiframe superresolution restoration, PhD Thesis, University of Notre Dame (2004)
29. Chaudhuri, S.: *Super Resolution Imaging*, Kluwer Academic Publishers (2000)
30. Teyssieux, D., Thiery, L., and Cretin, B.: Near-infrared thermography using a charge-coupled device camera: Application to microsystems, Rev. Scient. Instr. **78**, 034902 (2007)
31. Alaruri, S., et al.: Mapping the surface temperature of ceramic and superalloy turbine engine components using laser-induced fluorescence of thermographic phosphor, Optics and Lasers in Engineering **31**, 354–351 (1999)

Chapter 15
Near-Field Optical Microscopy in the Infrared Range

Yannick De Wilde, Paul-Arthur Lemoine, and Arthur Babuty

15.1 Introduction

The infrared covers the region of the electromagnetic spectrum with wave numbers (optical frequencies) in the range 13 000 to 10 cm^{-1}, which corresponds to the wavelength range from 780 nm to 1 mm. Infrared frequencies, in particular those in the mid-infrared ($2 < \lambda < 25$ μm), are especially attractive for probing materials, for they provide a way of identifying and quantifying their composition and structure, or probing their electronic and thermal properties. The absorption of infrared photons excites molecular vibrations and phonons. It is maximal at those frequencies associated with the vibrational modes, which depend on the kind of molecules and functional groups involved in the vibrations, or the crystal structure. So the frequencies of the absorption maxima form a spectral fingerprint of the material. Other factors such as the electron density and interactions between electrons, in semiconductors, metals, and superconductors, affect the dielectric properties at infrared frequencies, and are studied in the field of condensed matter. Regarding thermal imaging, infrared thermography is used to observe the temperature distribution at the surface of materials. The intensity of the thermal radiation emitted by a material depends on the temperature according to Planck's law. The temperature of a material can be determined by measuring the intensity of the infrared radiation it emits, once its emissivity is known.

Infrared spectroscopy and microscopy are widely used in research and industry. Instruments for making optical measurements in the infrared are generally based on the detection of propagating waves, collected and guided by conventional optical components, e.g., lenses and mirrors, into a detector located some distance from the sample. They are thus diffraction limited and for this reason unable to provide any information whatever concerning the optical properties of the sample on length scales shorter than about half the observation wavelength. The characteristic wavelengths of molecular vibrations are typically somewhere between a few micrometers and several tens of micrometers. The resolution in vibrational infrared spectroscopy

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 439
DOI 10.1007/978-3-642-04258-4_15, © Springer-Verlag Berlin Heidelberg 2009

is thus limited to at best a few micrometers. It is also at wavelengths of about 10 μm that the thermal energy radiated by materials at around room temperature is maximal. The spatial resolution in conventional infrared thermal microscopy is therefore also limited to a few micrometers.

To cope with recent developments in nanoscience and nanotechnology, new methods have been devised for investigating the optical properties of nanocomposites and nanostructured devices in the infrared, going beyond the diffraction limit. In order to get round this intrinsic limit in instruments like the conventional optical microscope, the only way is to probe the near field at the sample surface, examining those components of the electromagnetic field that vary on smaller length scales than the observation wavelength λ. To achieve this, the scanning near-field optical microscope (SNOM) [1] was invented shortly after the scanning tunneling microscope (STM) [2] and the atomic force microscope (AFM) [3]. In SNOM, a probe of sub-wavelength dimensions is scanned through the near-field region under piezoelectric control. The recordings of the near field at the location of the probe as it scans are used to acquire an optical image of the scanned region of the surface point by point, with a resolution that depends only on its size.

SNOM often uses a nanoaperture at the end of an optical fibre sharpened to a point in order to probe the near field. This technique is applied in particular to carry out optical studies in the visible or near infrared ($\lambda < 2$ μm), where conventional optical fibres are not very absorbent. Scattering SNOM (s-SNOM), on the other hand, uses light scattered by an AFM tip to form an image of the near field. With this apertureless approach, the electromagnetic waves do not have to be guided by an optical fibre, so it can operate in a broad spectral range, from the visible ($\lambda \sim 0.5$ μm) [4] to terahertz waves ($\lambda \sim 300$ μm) [5, 6]. S-SNOM does not involve guiding the waves through an aperture of size less than λ, and it generally has better resolution than aperture SNOM. The resolution in s-SNOM microscopy depends only on the size of the apex of the AFM tip used to scatter the near field. Infrared optical images can thus be produced for $\lambda \sim 10$ μm with resolutions of the order of around ten nanometers [7].

In this chapter, we explain the basic principles making it possible for near-field microscopy to reach a resolution well beyond the classical diffraction limit. Since s-SNOM is much more widely used than aperture SNOM in the infrared, we shall pay particular attention to this probe, describing how it works and showing by several recent applications how it can be used to study the optical properties of nanomaterials in the infrared, or to probe the electromagnetic field distribution over optoelectronic devices. Finally, we describe a new type of local probe directly inspired by s-SNOM, called the thermal radiation scanning tunneling microscope (TRSTM), able to detect the thermal emission produced by a sample surface in the mid-infrared with a resolution of the order of a hundred nanometers.

15.2 Resolution Limit in Conventional Microscopy

The image of a point source radiating at wavelength λ, formed using a conventional optical system, is always of finite size due to spatial filtering occurring when the electromagnetic waves propagate from the plane of the point source (object plane) to the plane of the image (image plane). A point in space can be characterised by a delta function, to which there corresponds an infinite spectrum of spatial frequencies k_x and k_y. In this section, we shall show that part of the spectrum satisfying $k_x^2 + k_y^2 > (2\pi/\lambda)^2$ is generally lost directly at the sample surface because it is associated with evanescent waves. Regarding the waves that propagate, some of them go in directions that are not compatible with entering the finite angular opening of the collecting optical system, producing further spatial filtering.

The finite size of the aperture in optical instruments detecting propagating electromagnetic waves leads to a cutoff in the angular distributions of wave vectors contributing to image formation, and produces diffraction [8]. In conventional microscopy, the effect of diffraction on a point source is to produce an image whose intensity distribution is given by a Bessel function, also known as an Airy pattern.

The Airy pattern has circular symmetry. There is a very intense central spot, the Airy disk, surrounded by concentric rings whose intensity decreases with the distance from the center. The radius R of the Airy disk, defined as the distance between the first zero of the Airy pattern and the center of the disk, is equal to $R = 1.22\lambda/2n\sin\theta$, where n is the optical index of the material in which the light emitted (or scattered) by the object propagates, and θ is the light gathering half-angle of the microscope objective. The product $n\sin\theta$ is called the numerical aperture (NA) of the microscope objective. Silica-based objectives adapted to the visible or near infrared can have a numerical aperture as high as 0.95–1.4 (depending on the value of n in the medium, which may be air with $n = 1$, or a liquid with higher index, such as oil, for which $n = 1.51$). Mirror objectives used in the mid-infrared generally have a smaller angular aperture, with NA equal to at best 0.5.

The limiting resolution in conventional microscopy is illustrated in Fig. 15.1. Two point sources located in the object plane produce Airy patterns that partially overlap in the image plane. When the distance ε between the point sources is small, the overlap between the Airy disks is such that their images cannot be individually resolved. The limiting resolution in conventional microscopy is usually expressed through the Rayleigh criterion. This expresses the idea that the resolution limit Δr of the microscope is reached when the distance ε between the point sources is such that the first minimum of the Airy pattern produced by the first source coincides with the center of the Airy disk associated with the second in the image plane. According to the theory of diffraction, the diffraction limit of a conventional microscope is given by

$$\Delta r = 1.22\frac{\lambda}{2n\sin\theta} = 1.22\frac{\lambda}{2\,\mathrm{NA}}\,. \tag{15.1}$$

An object of arbitrary shape and size can be considered as made up of an infinite set of point sources. The resolution Δr of a conventional microscope specifies the

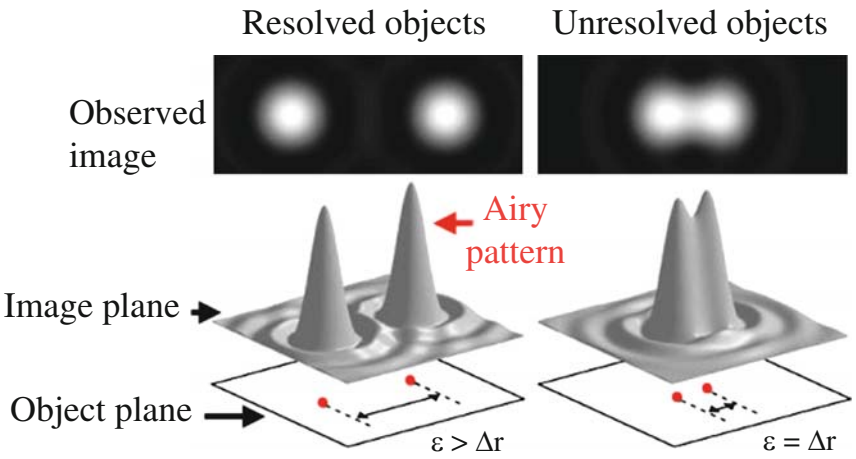

Fig. 15.1 Illustration of the Rayleigh criterion in conventional microscopy. Two point sources located in the object plane of the microscope produce Airy patterns in the image plane. When the distance ε between the point sources is less than $\Delta r = 1.22\lambda/2\mathrm{NA}$, the overlap of the Airy disks is too great to be able to resolve each of the points separately. The *upper part* of the figure represents the images seen with a microscope when the two point sources are separated by distances ε greater than or equal to Δr [9]

size of the smallest detail of the object that can be resolved by this microscope. For observations in the visible ($\lambda \approx 500$ nm), the best microscopes reach a resolution of the order of 250 nm. Observations made using conventional infrared microscopy at $\lambda \approx 10$ μm using a mirror objective (Cassegrain objective) are limited in resolution by diffraction at around 10 μm, given the smaller NA values for this type of objective (NA = 0.5 typically). In the infrared, it is therefore essential to find some way around the diffraction barrier in order to achieve nanometric resolutions.

There is another intrinsic limit to the resolution. It results from the evanescent nature of the waves associated with the high spatial frequencies of the object, which remain confined to the vicinity of the surface. As conventional microscopes operate solely by detecting propagating waves by means of a detector located at some distance greater than λ, these evanescent waves do not contribute to image formation, whatever the NA value of the optical system.

Consider the situation depicted in Fig. 15.2, in which a nanostructured object has a broad range of spatial frequencies, and assume that this object is illuminated by a plane wave of wave vector $k = 2\pi/\lambda$. Using the angular spectral representation of the electromagnetic fields [10], the exact solution of the wave propagation equation for the electric field at an arbitrary distance z above the sample surface is

$$\mathbf{E}(x,y,z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{\mathbf{E}}(k_x,k_y,0)\mathrm{e}^{\mathrm{i}(k_x x+k_y y+k_z z)}\mathrm{d}k_x\mathrm{d}k_y\,, \tag{15.2}$$

where k_z satisfies

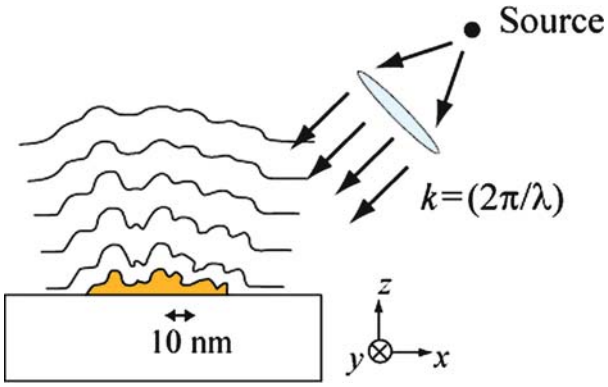

Fig. 15.2 The angular spectral representation allows exact solution of the wave equation, showing that the distance that separates the surface of a nanostructured object illuminated by a plane wave plays the role of a low-pass filter. The only Fourier components of the field in the plane of the object ($z = 0$) that are able to propagate over large distances ($\gg \lambda$) to an optical detector are those with transverse variations occurring on large scales with respect to λ, such that $k_x^2 + k_y^2 \leq k^2$. The high spatial frequencies, associated with small details of the object, produce evanescent contributions which remain confined to the sample surface

$$k_z = \sqrt{k^2 - k_x^2 - k_y^2}\,, \qquad \mathrm{Im}\, k_z \geq 0\,. \tag{15.3}$$

Here x, y are Cartesian coordinates defined in the object plane and k_x, k_y are the associated spatial frequencies.

The amplitude $\hat{\mathbf{E}}(k_x, k_y, 0)$ appearing in (15.2) is the Fourier transform of the original electric field in the object plane ($z = 0$). A similar expression can also be derived for the magnetic field. Equation (15.2) shows that $\mathbf{E}(x, y, z)$ can be described as a superposition of plane waves and evanescent waves whose amplitudes are given by the Fourier transform of the electric field in the plane of the sample, i.e., where $z = 0$. Indeed, according to (15.3), the component k_z of the wave vector can be either real or imaginary. The factor $\exp(\mathrm{i}k_z z)$ appearing in (15.2) is therefore either an oscillating function (propagating plane wave), or an exponentially decreasing function (evanescent wave), depending on the value of the spatial frequencies k_x and k_y. For a given pair of spatial frequencies (k_x, k_y), we have

$$\begin{aligned} &\mathrm{e}^{\mathrm{i}(k_x x + k_y y)} \mathrm{e}^{\mathrm{i}|k_z| z}\,, \qquad k_x^2 + k_y^2 \leq k^2 \qquad \text{(plane waves)}\,, \\ &\mathrm{e}^{\mathrm{i}(k_x x + k_y y)} \mathrm{e}^{-|k_z| z}\,, \qquad k_x^2 + k_y^2 > k^2 \qquad \text{(evanescent waves)}\,. \end{aligned} \tag{15.4}$$

Hence the field components associated with large details in the object plane ($k_x^2 + k_y^2 \leq k^2$) give rise to plane wave that can propagate to large distances and be measured using an optical detector. The small details ($k_x^2 + k_y^2 > k^2$) produce evanescent waves that remain confined to the surface of the object. These cannot be detected far from the object unless some specific strategy is implemented to convert them into propagating waves that could then be picked up by a detector located in the far

field. According to (15.3) and (15.4), the confinement will increase with the value of the spatial frequency ($k_x^2 + k_y^2 \gg k^2$), which corresponds to spatial variations of the field at the surface of the object occurring at length scales much shorter than λ. It is precisely these spatial frequencies that are needed to form an optical image of the object with sub-wavelength resolution.

15.3 Near-Field Microscopy

15.3.1 Basic Idea

According to what has just been shown, we may conclude that, in electromagnetism, the distance to the surface of an object plays the role of a low-pass filter [11]. When an object is observed at a wavelength λ, such that $k = 2\pi/\lambda$, all the spatial frequencies of the electromagnetic field associated with the object can be found in the plane of the object ($z = 0$). However, the high spatial frequencies, such that $k_x^2 + k_y^2 > k^2$, are filtered during propagation, in such a way that the information relating to spatial variations of the electromagnetic field that are small compared with λ will be lost. Structures with transverse dimensions

$$\delta < \frac{1}{k} \left(= \frac{\lambda}{2\pi} \right), \tag{15.5}$$

produce evanescent waves that remain confined to the surface of the sample. The near field is defined as the region very close to the sample surface in which are confined the Fourier components of the electromagnetic field associated with high spatial frequencies giving rise to evanescent waves. The far field is the region of space in which waves associated with low spatial frequencies of the electromagnetic field in the object plane propagate far from the object.

In order to obtain optical images of a nanostructured sample with sub-wavelength resolution, the information contained in the near field must be collected. In a SNOM, this is made possible by using a probe with dimensions very much smaller than the wavelength. The probe scans the sample surface in the near-field region, whilst maintaining a maximal distance of a few nanometers above the surface. Figure 15.3 shows how this local optical probe is used to convert the near field at the probe position into a propagating field, thus making it detectable by means of an optical detector placed in the far-field region. In order to carry out controlled displacements on the nanoscale at the sample surface, the local optical probe is generally mounted on an AFM cantilever. By recording the near field at each new position of the probe as it scans over the sample surface, an optical image of the scanned region is built up point by point, with a resolution that depends only on the size of the probe. This image of the optical near field, which we shall call the SNOM image, is obtained at the same time as the topographic AFM image.

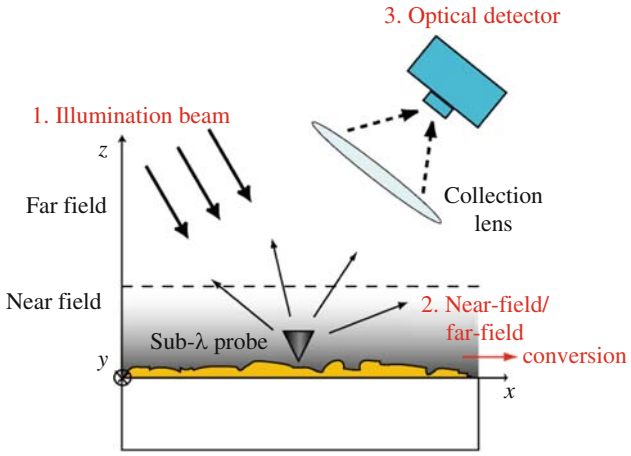

Fig. 15.3 Principle of the near-field optical microscope

There are different types of SNOM. The most widely used are aperture SNOM and scattering SNOM. For infrared studies, the latter are generally favoured. Note also a third type of SNOM used in the visible and near infrared, viz., SNOM with active fluorescent tips. These are described in detail in Chap. 17 and we shall not discuss them further here.

15.3.2 Aperture SNOM

The original idea for the near-field optical microscope is attributed to the physicist Synge [12]. In 1928, he suggested using an aperture with sub-wavelength dimensions in an opaque metal screen to illuminate the sample surface locally, as shown schematically in Fig. 15.4. By scanning the aperture above the sample surface at a distance less than λ, only a region of the surface with size less than λ should be illuminated owing to the confinement of the electromagnetic field close to the aperture, which thus constitutes a nanoscale light source. When the sample is composed of nanoscale objects arranged on a transparent substrate, Synge predicted that the nano-objects would scatter the field produced by the nanosource, and that they would convert the evanescent fields into propagating fields that could be detected by means of a detector located in the far field. In this thought experiment, the idea was to produce an optical image of the scanned region by measuring the intensity of the optical signal collected across the sample at each point scanned by the nanoscale aperture, with a resolution determined solely by the diameter of the nanoscale aperture and by its distance from the sample surface.

However, it was not until half a century later that the physicist Pohl put Synge's idea into practice [12], shortly after the invention of the STM [2] and the AFM [3].

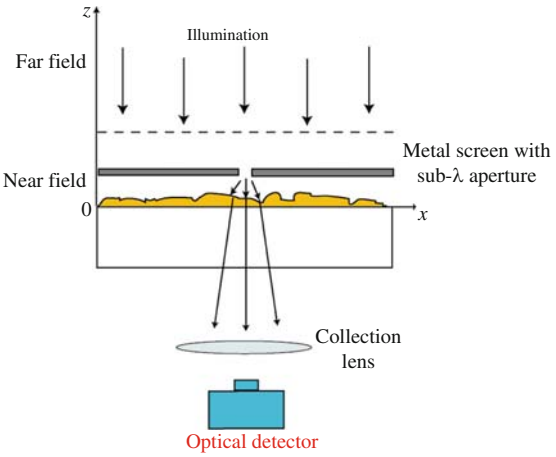

Fig. 15.4 Principle of aperture SNOM as proposed by Synge in 1928 [12]. An optical source of dimension much less than the wavelength λ is produced by illuminating through a metal screen containing a hole of nanometric dimensions

Most aperture SNOMs use an optical fibre sharpened to a point to guide the light from the near to the far field. Applications of aperture SNOM use silica fibres and concern mainly the visible and near infrared wavelength regions. The tips of these fibres can be made by chemical etching with an HF solution [13, 14] and by hot drawing, following a procedure originally developed to make micropipettes [15]. To make the nanoscale aperture, the silica tip is generally coated with metal before truncating it at the end with a focused ion beam (FIB). This produces an aperture of a few tens of nanometers at the center of a flat region, corresponding to Synge's original idea of producing an opaque screen containing a hole of sub-wavelength dimensions [16].

The most widely used method for controlling the position of this aperture on the sample surface is to fix the silica fibre along one arm of a quartz tuning fork. By exciting the mechanical resonance of the tuning fork, e.g., using a piezoceramic, a lateral oscillation of the silica tip is generated above the sample surface. The amplitude of this oscillatory motion is measured using the piezoelectric signal arising at the terminals of the quartz tuning fork. When the tip is only separated from the surface by a few tens of nanometers, friction forces come into play between the tip and surface. They reduce the amplitude of the lateral oscillation of the tip, and increasingly so as the tip approaches the surface. This dependence of the damping effect on the distance is used to servocontrol the separation between tip and surface in such a way as to hold it constant during scanning [17].

Beyond $\lambda = 3$ μm, the optical transmission of silica becomes rather poor, making it unusable for mid-infrared applications of near-field microscopy. To get around this problem, several groups set to work on aperture SNOM using fibres specially designed for the mid-infrared. These are chalcogenide glass fibres, e.g., As_xSe_{1-x}, or fibres made from AgClBr, which have been developed recently [18–20]. However,

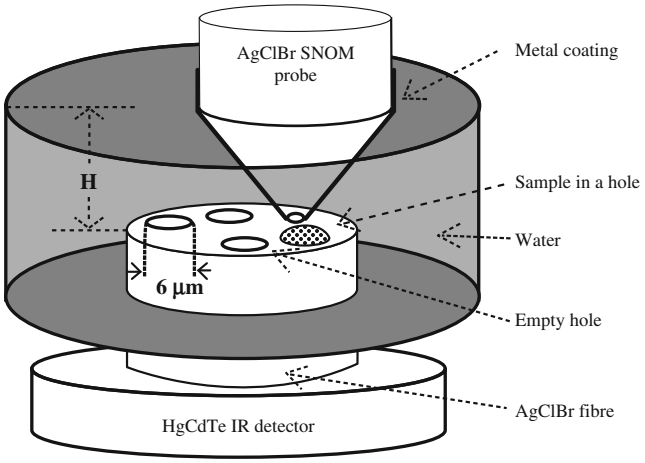

Fig. 15.5 Aperture SNOM operating in the mid-infrared at $\lambda = 10.6$ μm [23]. In this setup, the sample is totally immersed in water and mounted on an AgClBr fibre. This guides the light transmitted through the sample to an infrared (HgCdTe) detector. A second AgClBr fibre ending with a nanoscale aperture is partially immersed in the water and scans the sample surface. This fibre guides the infrared light produced by a CO_2 laser ($\lambda = 10.6$ μm) and locally illuminates the sample surface, thus removing the problem raised by the low transmission of the water layer covering the sample. Adapted from [23]

it is not easy to use these fibres in an aperture SNOM. They have diameters of the order of 0.5–1 mm, which is 10 times the diameter of silica fibres. They are also around 500 times heavier [20]. As a result, quartz tuning forks cannot be used and an alternative solution must be found for holding the fibre and controlling its position when it oscillates above the surface. Furthermore, an exact calculation of the transmission T of an aperture with sub-wavelength diameter d in a perfectly conducting and infinitely thin metal film predicts the following dependence [21,22]:

$$T \propto \frac{d^4}{\lambda^4} \,. \tag{15.6}$$

The light intensity transmitted through the nanoscale aperture of a SNOM operating in the mid-infrared is therefore expected to be very low compared with what can be obtained in the visible. This effect can be partially compensated by increasing the diameter of the aperture, but to the detriment of the resolution that can be obtained from the instrument, and also by using intense enough sources for the optical signal transmitted by the aperture to reach a detectable level.

Despite the difficulties involved in making an aperture SNOM that can operate in the mid-infrared, in some circumstances it may be useful to have a nanosource guiding the light right to the sample surface in a low-absorption fibre. The main aim of applications stimulating the development of aperture SNOM in the mid-infrared is to carry out studies of biological samples immersed in a highly absorbent medium such as water, which make it impossible to illuminate from the outside. When the

sample is immersed in a water layer several hundred micrometers thick, there is a considerable advantage in placing the nanoscale infrared light source in direct contact with the sample, so as to minimise absorption losses in the water. Figure 15.5 shows an example of an experimental device designed with this in mind [23].

15.3.3 Apertureless or Scattering SNOM

Principle of Scattering SNOM

As with aperture SNOM, scattering SNOM can detect the near field and reach subwavelength resolutions. This method, also known as apertureless SNOM, exploits the fact that a particle smaller than the wavelength of an external electromagnetic field will behave as a dipole, radiating a field with intensity proportional to the field intensity at the position of the dipole. Figure 15.6 shows a conventional microscope image, obtained in the far field, of gold nanospheres ($\phi = 100$ nm) dispersed on a glass slide and illuminated with white light ($\lambda \sim 500$ nm). The particles in this image appear as diffraction-limited spots with diameter $d_s \sim \lambda/\mathrm{NA}$. Each of them plays the role of a nanoscale antenna that radiates a field of intensity I proportional to the intensity I_0 of the field where it happens to be located.

Figure 15.7 illustrates the basic operation of apertureless SNOM. Consider a nanostructured sample, illuminated in reflection or transmission, while at the surface a nanosphere that is much smaller than the illumination wavelength scatters the near field which is present at its location. The light scattered by the nanosphere is gathered by a collection lens (or objective) and then focused on an optical monodetector, as shown in Fig. 15.7a. Now if it were possible to displace the nanosphere in

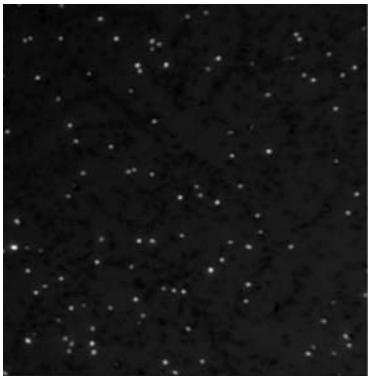

Fig. 15.6 Image of gold nanospheres ($\phi \sim 100$ nm) on a glass slide, obtained in white light ($\lambda \sim 500$ nm) with a conventional microscope (NA = 0.95). Due to the diffraction limit, the nanoparticles cannot be resolved, and appear as bright spots of diameter around 500 nm. However, the field intensity they scatter is proportional to the local field intensity

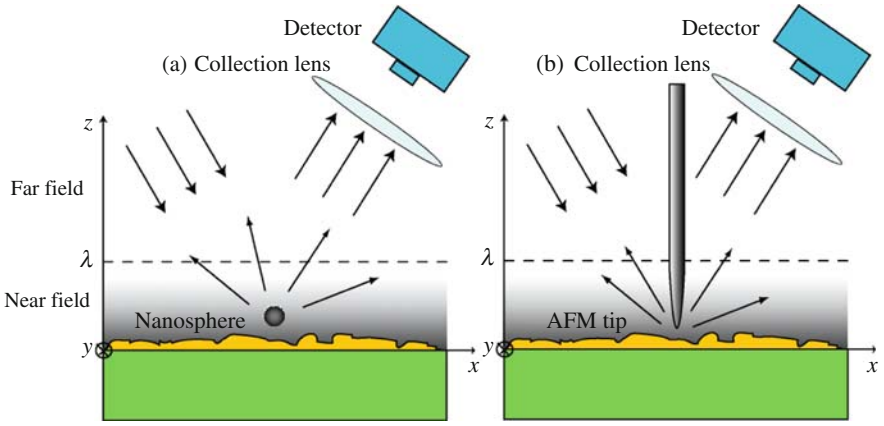

Fig. 15.7 (**a**) The idea in scattering SNOM (s-SNOM) is to scan the near-field region with a nanoscatterer, which acts rather like a nanoscale antenna to detect the near field locally. The scattered field is gathered by a collection lens and focused on a monodetector. (**b**) In practice, the metal tip of an atomic force microscope is used as nanoscatterer. Its position during scanning can be controlled to within the nanometer

a controlled way so as to scan the sample surface in nanometric steps, a recording of the scattered field intensity as a function of position would build up, point by point, an optical image of the scanned region. The optical resolution expected in such an experiment would depend only on the diameter of the nanosphere used as nanoscale antenna to pick up the local near field. In particular, it would be independent of the illumination wavelength. So the diffraction limit could be overcome if there were some way of displacing a nanosized scatterer through the near-field region.

In order to implement this idea, an atomic force microscope (AFM) with a metal tip is generally used, as shown in Fig. 15.7b. The end of the AFM tip, or apex, is a local scatterer of the near field whose size is determined by the radius of curvature of the apex. Its position on the sample surface is controlled using a piezoelectric system acting in the three orthogonal directions x, y, and z. Under computer control, this system can scan up to several tens of micrometers in the lateral directions (denoted by x and y in Fig. 15.7) with nanometric precision. An electronic servosystem acting on the z piezo maintains a constant value of the average tip–surface separation. A topographic image of the surface is obtained by recording the servo signal as a function of the lateral position of the tip.

Interpreting the Contrast. The Image Method

The contrast in s-SNOM images is a dielectric contrast. The simplest mode for explaining this starts by assuming that the end of the tip forms a spherical dipole **p** of radius r and polarisability

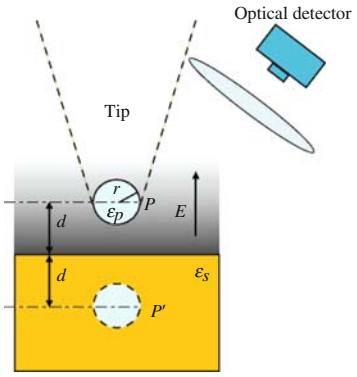

Fig. 15.8 Coupling of the probe dipole (*continuous line*) with the image dipole

$$\alpha = 4\pi r^3 \frac{\varepsilon_\mathrm{p} - 1}{\varepsilon_\mathrm{p} + 2} \,, \tag{15.7}$$

where ε_p is the dielectric constant of the tip at the relevant optical frequency. This probe dipole interacts with its image $\mathbf{p}'$ in the surface, as shown in Fig. 15.8. When r and the distance d separating the tip and surface are very small compared with λ, phase shift effects due to propagation are negligible, and we may appeal to the electrostatic limit [24]. In the near-field region, the probe dipole feels an electric field $\mathbf{E}$ that we shall assume here to be perpendicular to the surface. It polarises with a dipole moment $\mathbf{p} = \alpha\mathbf{E}$. For its part the image dipole has a dipole moment $\mathbf{p}' = \beta\mathbf{p}$, where

$$\beta = \frac{\varepsilon_\mathrm{s} - 1}{\varepsilon_\mathrm{s} + 1} \,.$$

It is of course located at a distance $2d$ from the probe dipole.

The electric field produced by this image dipole modifies the field at the position of the probe dipole by an amount

$$E_\mathrm{image} = \frac{p'}{2\pi(2d)^3} \,. \tag{15.8}$$

The probe dipole feels the superposition of the two fields E and E_image. Hence p is modified in such a way that

$$p = \alpha(E + E_\mathrm{image}) = \alpha\left(E + \frac{\beta p}{16\pi d^3}\right) \,, \tag{15.9}$$

and consequently,

$$p = \frac{\alpha E}{1 - \alpha\beta/16\pi d^3} \,. \tag{15.10}$$

For the coupled system $\mathbf{p}_\mathrm{T} = \mathbf{p} + \mathbf{p}'$, the effective polarisability α_eff defined by $\mathbf{p}_\mathrm{T} = \alpha_\mathrm{eff}\mathbf{E}$ depends on the dielectric properties of the tip, but also on the dielectric

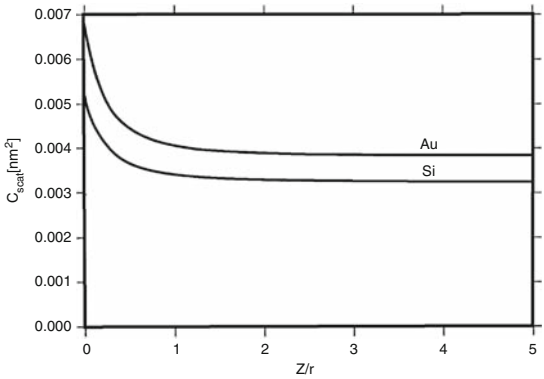

Fig. 15.9 Scattering cross-section of a gold sphere of radius 30 nm coupled with its image in a gold or silicon surface. From [24]

properties of the material making up the surface directly below the tip:

$$\alpha_{\text{eff}} = \frac{\alpha(1+\beta)}{1-\alpha\beta/16\pi d^3} \,. \tag{15.11}$$

Having found the effective polarisability α_{eff} using this quasi-electrostatic model, the Mie theory of light scattering by a spherical particle of size much smaller than unity (the Rayleigh limit) can be applied [25] to calculate the scattering cross-section C_{scat}:

$$C_{\text{scat}} = \frac{k^4}{6\pi}\left|\alpha_{\text{eff}}\right|^2 \,. \tag{15.12}$$

Recall that, in the theory of light scattering, the scattering cross-section has units of area [m^2]. It is used to calculate the total power P_{scat} [W] scattered into the far field, over the whole solid angle, by the scatterer as a function of the incident light intensity I_0 [W/m^2] through the relation $P_{\text{scat}} = C_{\text{scat}} I_0$. Figure 15.9 shows the behaviour of the scattering cross-section C_{scat} of a gold (Au) probe sphere coupled with its image as a function of the distance to an Au or Si surface, calculated using (15.11) and (15.12). The coupling between the two dipoles induces a significant increase in C_{scat} when the distance between the probe sphere and the surface is of order r [24].

In a completely general situation, the dielectric constants ε_{p} and ε_{s} of the probe and sample, respectively, can be complex valued, whence α_{eff} will also be complex valued. The relation between the incident field E and the field E_{scat} scattered into the far field is

$$E_{\text{scat}} \propto \alpha_{\text{eff}} E = \left|\alpha_{\text{eff}}\right| \mathrm{e}^{\mathrm{i}\phi_{\text{scat}}} E \,, \qquad \phi_{\text{scat}} = \text{Arg}(\alpha_{\text{eff}}) \,. \tag{15.13}$$

This means that E_{scat} can acquire an extra phase ϕ_{scat}, depending on whether or not α_{eff} has an imaginary part.

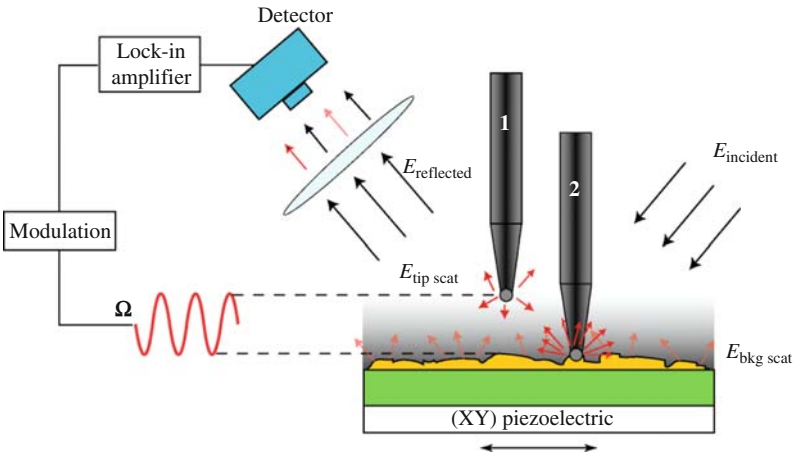

Fig. 15.10 Recording the s-SNOM signal by tapping mode AFM and a lock-in amplifier method. Note that a background is usually superposed on the field scattered by the tip, caused by diffuse or specular reflection from the sample surface

Extracting the Signal from the Near Field

The main advantage of s-SNOM over aperture SNOM is that no wave guides are needed, which means that it can operate potentially at any wavelength from the visible, through the infrared to terahertz waves. The price to pay for this is that, very often in measurements made with s-SNOM, a significant background signal reaches the detector at the same time as the light scattered by the tip. This background is due to specular or diffuse reflection occurring at the sample surface and on the upper regions of the probe itself. It contains information of no value that is likely to cause artifacts in the images. Ways of suppressing this background have been devised, extracting the contribution from scattering by the tip apex, which reveals the near-field interaction between the tip and sample.

One way of eliminating these unwanted contributions is to oscillate the tip normally to the sample surface using tapping mode AFM at a frequency Ω, and demodulating the optical signal at Ω or one of its harmonics Ω_2, Ω_3, as shown in Fig. 15.10 [24, 26]. The key feature of this method is the nonlinear dependence of α_{eff} and the fact that the near-field interaction falls off over distances comparable with the tip radius ($r \ll \lambda$), whereas the background varies over distances comparable with λ. If the tip oscillates at frequency Ω, with amplitude $\Delta d \ll \lambda$, the near-field interactions produce scattered light at this frequency and its harmonics ($\Omega_n = n\Omega$, $n = 2, 3, \ldots$). In contrast, since the background has slow spatial variations, it contributes essentially through a continuous term or possibly through lower harmonic terms. The relative importance of these terms compared with the contribution from the near-field scattered by the tip apex depends on λ and Δd.

The radius r of the s-SNOM tip apex is generally in the range 10–100 nm. In the scattering sphere approximation, C_{scat} is very small and the expected optical

signals are thus very weak. The small size of the scatterer which contributes to the near-field signal is in practice partly compensated by the effect of local field enhancement under the tip. The origin of this enhancement effect is similar to what happens in electrostatics in the vicinity of sharply pointed objects, and which is put to use in lightning conductors [27,28]. Although difficult to quantify experimentally, this effect can lead to an antenna gain of several orders of magnitude for the field component parallel to the principal axis of the tip [27, 28].

Furthermore, s-SNOM systems with laser sources often use either homodyne or heterodyne interferometric setups [29]. These arrangements can increase the level of the optical signals to be detected, and aim to separate information relating to the phase and the amplitude of the field scattered by the tip. Figure 15.11 illustrates the principle of homodyne (Fig. 15.11a) and heterodyne (Fig. 15.11b) detection. A detailed description of these techniques applied to s-SNOM can be found in [30].

In homodyne detection, the intensity of the signal at the detector results from interference between the field E_{scat} scattered by the tip and a reference field E_{ref} which oscillates at the same optical frequency ω. These fields can be written in the form

$$E_{\mathrm{ref}} = E_{0\mathrm{ref}} \mathrm{e}^{\mathrm{i}(\omega t + \phi_{\mathrm{ref}})} , \qquad E_{\mathrm{scat}} = E_{0\mathrm{scat}} \mathrm{e}^{\mathrm{i}(\omega t + \phi_{\mathrm{scat}})} , \tag{15.14}$$

where E_0 and ϕ are the corresponding amplitude and phase of the fields. The intensity at the detector is

$$I = \left|E_{\mathrm{ref}}\right|^2 + 2\left|E_{\mathrm{ref}}\right|\left|E_{\mathrm{scat}}\right| \cos(\phi_{\mathrm{scat}} - \phi_{\mathrm{ref}}) + \left|E_{\mathrm{scat}}\right|^2 . \tag{15.15}$$

Only the second and third terms of this expression are modulated by the tip and hence able to produce a signal that can be detected by a lock-in amplifier technique (which demodulates the optical signal at Ω or a higher harmonic). Furthermore, generally $\left|E_{\mathrm{ref}}\right| \gg \left|E_{\mathrm{scat}}\right|$, so it is mainly the term $2\left|E_{\mathrm{ref}}\right|\left|E_{\mathrm{scat}}\right| \cos(\phi_{\mathrm{scat}} - \phi_{\mathrm{ref}})$ that contributes to the signal. The field $\left|E_{\mathrm{ref}}\right|$ thus has the effect of amplifying the field scattered by the tip, making it more easily detectable. Many s-SNOM setups simply use the background signal as reference (see Fig. 15.10). Scattering SNOM setups equipped with an arm for the reference beam (see Fig. 15.11a) include a moving mirror allowing adjustment of the phase ϕ_{ref}. The mirror is initially positioned in such a way as to optimise the optical signal, so that $2\left|E_{\mathrm{ref}}\right|\left|E_{\mathrm{scat}}\right| \cos\phi_{\mathrm{scat}}$ is measured by a lock-in system. The mirror is then shifted by $\lambda/8$ to produce a path difference of $\lambda/4$ in the interferometer and hence a phase difference of 90° relative to the first measurement, so that $2\left|E_{\mathrm{ref}}\right|\left|E_{\mathrm{scat}}\right| \sin\phi_{\mathrm{scat}}$ is then measured. With these two measurements, the amplitude and phase of the field scattered by the tip can be extracted independently of one another. Note that when the tip is illuminated by a purely evanescent field and there is no reference beam, the intensity $\left|E_{\mathrm{scat}}\right|^2$ of the field scattered by the tip is directly accessible because only the third term of (15.15) contributes to the signal.

In a heterodyne detection setup, a very small shift (a few tens of MHz at the most) is introduced between the optical frequencies of the reference and scattered fields, so that a beat appears in the intensity at the optical detector:

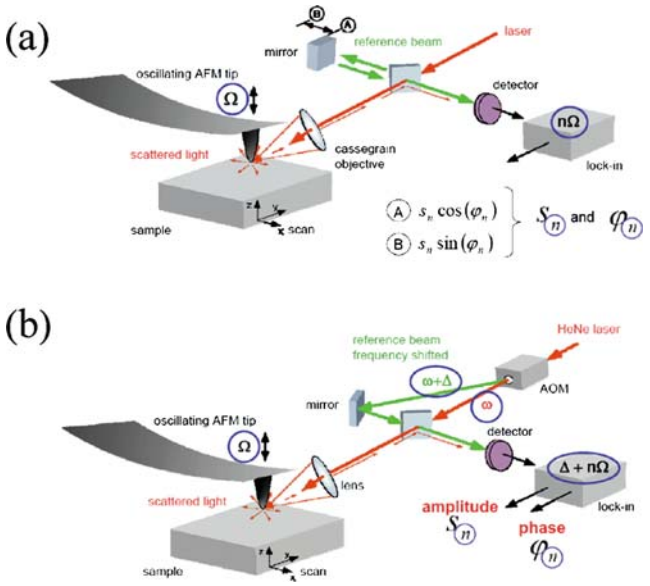

Fig. 15.11 (**a**) Homodyne and (**b**) heterodyne detection. Taken from Fig. 1 of [29]

$$I = \left|E_{\text{ref}}\right|^2 + 2\left|E_{\text{ref}}\right|\left|E_{\text{scat}}\right|\cos(\Delta\omega t + \phi_{\text{scat}} - \phi_{\text{ref}}) + \left|E_{\text{scat}}\right|^2 . \tag{15.16}$$

Taking into account the modulation of $\left|E_{\text{scat}}\right|$ due to the vertical oscillation of the tip, the amplitude and phase of the field scattered by the tip at the fundamental oscillation frequency of the tip, or one of its harmonics, can be extracted in a single measurement by demodulating the optical signal at the frequency $\Delta\omega + n\Omega$, where $n = 1, 2, 3, \ldots$, using a two-channel (amplitude and phase) lock-in amplifier system.

Applications in the Infrared

Mid-infrared s-SNOM has proven fruitful in many areas of research such as nanomaterials [31, 32], polymers [33], biology [34], semiconductors [35, 36], plasmonics [37, 38], and optoelectronics [39, 40]. Most research involves dielectric contrast mapping on samples with surfaces displaying sub-wavelength spatial variations in their crystal structure [41], chemical nature or composition [31], or doping in the case of semiconductors [35, 36]. Another important field of applications for infrared s-SNOM aims to visualise the electromagnetic field distribution on plasmonic structures [37, 38] or working laser devices [39, 40]. Finally, recent studies have shown that s-SNOM can be useful in condensed matter physics, demonstrating the relationship between nanoscale optical properties and electronic correlations [42].

Figure 15.12 gives an example of experimental studies of the dielectric contrast between two materials [31]. The sample is an Au film through which some regions

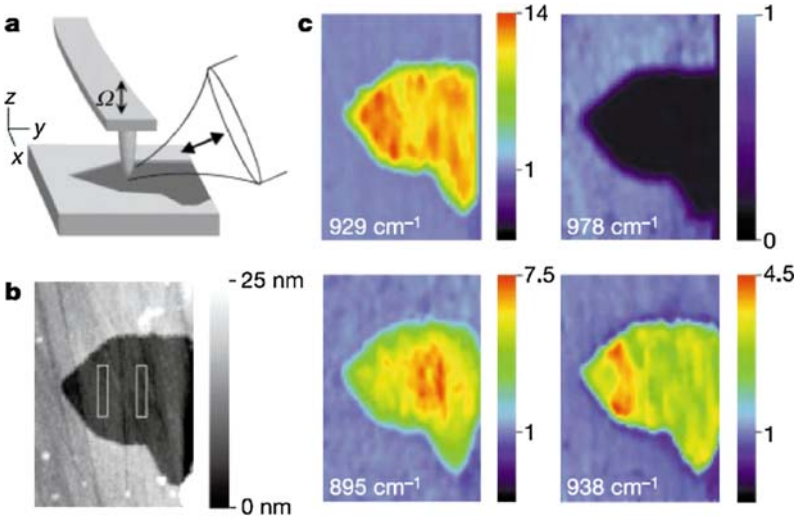

Fig. 15.12 (**a**) An infrared s-SNOM has been used to study a sample comprising an Au film with SiC islands. (**b**) AFM image of the region studied, and (**c**) s-SNOM images taken on this same region. Contrast variations between the two materials at different frequencies show good agreement with the quasi-electrostatic model describing the tip as a spherical dipole coupled with its image. Taken from [31]

of the SiC substrate of submicron dimensions are visible. A series of s-SNOM images were obtained on the same region of the sample at different optical frequencies close to 1 000 cm^{-1} ($\lambda = 10\ \mu\mathrm{m}$). The source was a CO_2 laser with selectable emission lines. The point here is that SiC has a polariton resonance in the range of optical frequencies accessible to the CO_2 laser lines. Neither the tip nor the Au have a resonance in this region of the electromagnetic spectrum. At the optical frequency with $\mathrm{Re}(\varepsilon_{\mathrm{SiC}}) = -1$, the polarisability β of the surface becomes large so that, according to the image dipole method, the scattering cross-section C_{scat} also becomes large. It is near these frequencies that the s-SNOM signal is dominant on the SiC in the experiments (see Fig. 15.12). The dependence of the contrast between the SiC and the Au on the optical frequency in the vicinity of the resonant peak of C_{scat} has been studied systematically, confirming the relevance of the image dipole method for describing the tip–surface interaction.

A second application of s-SNOM is shown in Fig. 15.13, where it is used to study the field distribution produced near a nanoscale optical antenna at the surface of a working quantum cascade laser [40]. In these electro-optical devices, the active region made from semiconducting heterostructures is sandwiched between two metal electrodes. Applying a voltage across the two electrodes gives rise to laser emission anywhere between the mid-infrared and terahertz frequencies. In some quantum cascade lasers, the cavity surface hosts an evanescent field whose structure reflects that of the modes in the laser cavity. These evanescent fields, which are in principle undetectable, have been directly observed using s-SNOM [39]. In the case shown in

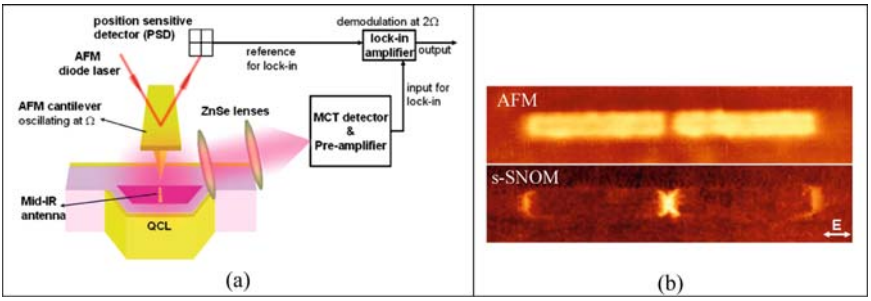

Fig. 15.13 (**a**) An infrared s-SNOM has been used to investigate a quantum cascade laser emitting at $\lambda = 7$ μm. (**b**) The device is equipped with an optical antenna in the form of two Au segments. When the antenna is suitably sized in relation to the emission wavelength, a very localised field arises between the two segments, and this field can be visualised using the s-SNOM. Length of probed region 3 μm. Taken from [40]

Fig. 15.13, a half-wave optical antenna made from two Au segments was deposited on one facet of the Fabry–Pérot cavity of a quantum cascade laser. The s-SNOM image shows that the effect of this antenna is to focus the light in a region of size smaller than 100 nm situated between the two Au segments. Given the emission wavelength of the laser, viz., $\lambda = 7$ μm, it is essential to use s-SNOM here in order to be able to resolve the spatial distribution of the field around the antenna.

15.4 Thermal Radiation STM

15.4.1 Introduction

Several kinds of scanning local probes have been developed over the last few years to probe the thermal properties of samples at the nanoscale. These local probes generally use a thermoresistive element or a miniaturised thermocouple placed in contact with the sample surface [43, 44]. The thermal nanoprobe can be used as a local temperature probe, or sometimes also as a local heat source. Other probes carrying a fluorescent active tip in contact with the sample exploit the temperature dependence of fluorescent emission to produce a temperature map [45].

When there is no external source and the temperature is not zero, thermal vibration of the charges in the materials leads to thermal radiation. For a black body, the power radiated by this internal source obeys Planck's law. The relation between the temperature T [K] and the wavelength $\lambda_{\max}$ [μm] of the emission maximum of the black body is given by

$$\lambda_{\max} = \frac{2898}{T} .$$

A material near room temperature is thus expected to emit mainly in the mid-infrared, around 10 μm. Far-field images of the infrared thermal radiation can be obtained by coupling an infrared camera with an infrared microscope. However, the diffraction limit as specified by the Rayleigh criterion means that the best possible resolution achievable in infrared thermal radiation images will be around ten micrometers. The optical near field is therefore the only opportunity for probing thermal radiation at the nanoscale.

Until very recently, it was commonly accepted that a SNOM, with or without aperture, required three essential ingredients:

- An external source, usually a laser, which illuminates the sample and produces a field at the sample surface.
- A scanning probe of sub-wavelength dimensions, i.e., a nanoscale aperture or scattering tip, able to pick up the near field locally and propagate it to a detector.
- An optical detector placed in the far-field region.

In this section, we shall describe a new near-field imaging device in which one of the three fundamental ingredients, namely the external source, has been eliminated. The thermal radiation scanning tunnelling microscope (TRSTM) is an optical near-field imaging device of the s-SNOM type which operates without any kind of external source. Based on the scattering of thermal radiation by means of a metal AFM tip, it can probe the local electromagnetic density of states, rather like a 'photon STM'. Indeed, the name TRSTM is due to the many analogies between this device and the STM based on the electron tunneling effect.

15.4.2 TRSTM Setup and Operation

The setup for TRSTM uses an infrared s-SNOM arrangement modified to be able to heat the sample up to around 200°C, but without overheating the sensitive parts of the equipment, such as the piezoelectrics and various bondings. Figure 15.14 is a schematic view [46]. It uses the same detection optics as an infrared s-SNOM. The sample is mounted on a copper stage whose temperature can be varied by the Joule effect using a heating resistance. The heater replaces the laser source here, stimulating thermal radiation in the sample. In experiments, the tungsten tip, mounted on a quartz tuning fork, oscillates normally to the sample surface at the frequency Ω, moving through the near-field region and scattering the thermal radiation periodically. This periodic contribution, due to the small perturbation produced by the

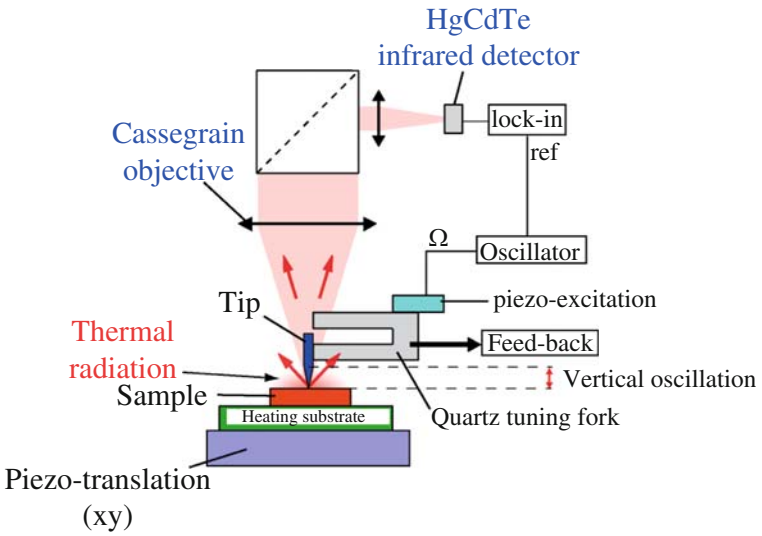

Fig. 15.14 Principle of the thermal radiation STM (TRSTM)

tip, is extracted from the continuous background by means of a lock-in amplifier which is connected at the output of the infrared detector. Given the weak signals expected, the setup aims to optimise the solid angle of collection of the photons scattered by the tip, and uses an HgCdTe infrared detector cooled with liquid nitrogen, with a detectivity D^* of around 4×10^{10} cm $\mathrm{Hz}^{1/2}\mathrm{W}^{-1}$ and spectral detection range ($> 50\%$ of the maximal detectivity) between 7 and 12 μm. Signals measured with the TRSTM are usually a few tens of pW, which is 3 or 4 orders of magnitude weaker than those detected with a near-field microscope using an infrared laser source.

The sample is scanned under the tip at the sample surface. The field scattered by the tip is detected at each point during the scan to build up an image of the near field of the thermal radiation point by point. The resolution is determined by the radius of curvature of the tip. A topographic image of the sample is also built up by recording the feedback signal during scanning.

15.4.3 First Example Application of TRSTM

The first near-field studies of infrared thermal radiation using the TRSTM were mainly concerned with gold nanostructures deposited on a silicon carbide (SiC) substrate. An image of one of these islands is shown in Fig. 15.15a. It is a gold stripe 1 mm long, with varying width between 10 and 35 μm. Figure 15.15b shows a series of AFM images and infrared thermal near-field images recorded at the same

time, obtained at a temperature of 170°C on regions of the gold stripe of different widths.

An interference filter centered at a wavelength of 10.9 μm (full width at half-maximum = 1 μm) is placed in front of the infrared detector for spectral selection of the energy of photons contributing to image formation. In the TRSTM images, the two materials can be clearly distinguished, with a stronger signal from the gold structure. The resolution achieved in the infrared near-field images of the thermal radiation is of the order of 100 nm, a factor of 100 better than the resolution that could be reached in conventional infrared thermal microscopy.

The contrast observed in Fig. 15.15b (top) between the gold and the SiC shows the ability of TRSTM to measure differences in the near-field thermal emission between two materials at the same temperature. Surprisingly, the TRSTM images exhibit fringes across the gold stripe, the number of which increases with the width of the stripe. These stationary wave patterns are the signature of coherence effects in the near-field thermal radiation close to the gold surface.

By analogy with the way an STM measures the local density of electronic states via the tunnel current, it is also possible to define an electromagnetic local density of states (EM-LDOS) $\rho(\mathbf{r},\omega)$ for photons [47]. This has the property that $\rho(\mathbf{r},\omega)\mathrm{d}\mathbf{r}$ gives the probability of finding a photon of energy $\hbar\omega$ in a volume d$\mathbf{r}$ centered at $\mathbf{r}$. At thermodynamic equilibrium, the local electromagnetic energy density $U(\mathbf{r},\omega)$ is given by the product of the EM-LDOS times the energy $\hbar\omega$ of the state times the average number of photons (Bose–Einstein distribution):

$$U(\mathbf{r},\omega) = \rho(\mathbf{r},\omega)\hbar\omega\frac{1}{\exp(\hbar\omega/kT)-1}\,. \tag{15.17}$$

In vacuum, the EM-LDOS is given by

$$\rho(r,\omega) = \frac{\omega^3}{\pi^2 c^3}\,,$$

i.e., it is a homogeneous and isotropic quantity. Combining with (15.17), this leads to the expression for the electromagnetic energy of a black body.

Recent theoretical [47–49] and experimental [50] work has shown that the EM-LDOS can exhibit significant variations in the vicinity of plane surfaces under the influence of resonant surface states, e.g., surface plasmons, phonon polaritons [47, 48], or in the presence of photonic nanostructures [49,50]. In order to probe the EM-LDOS, all modes of the system under investigation need to be populated, something that cannot be done as it is in s-SNOM by directional and polarised illumination, where the field at the sample surface is produced by a laser beam.

With TRSTM, the thermal excitation of the radiation ensures that all possible electromagnetic modes at the sample surface are populated in accordance with photon statistics. In this case, the intensity of the radiation scattered by the tip provides a direct measurement of the photon local density of states (EM-LDOS) at the position of the tip and projected along the tip axis [47]. This conclusion is supported by comparing TRSTM measurements made for different widths of the gold stripe

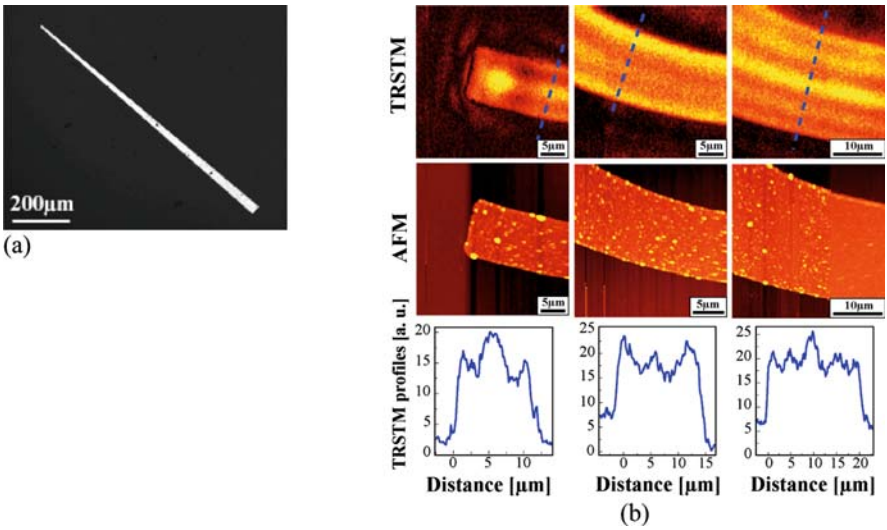

Fig. 15.15 (**a**) Photograph of a gold stripe of variable width on an SiC substrate. (**b**) *Top*: TRSTM images of the sample heated to 170°C, taken at different points of the structure. *Center*: Corresponding topographic images. *Bottom*: Profile of the top images, detailing the fringe structure

and theoretical calculations of the EM-LDOS projected along the tip axis, carried out by Boris Gralak at the Institut Fresnel (Marseilles, France) [46]. This projected EM-LDOS is the one that would describe the spontaneous emission rate of a dipole oriented along the tip axis.

The TRSTM measurements must take into account the extension of the scattering tip in the direction normal to the sample surface, which tends to average the EM-LDOS over several micrometers along the tip axis. Images obtained by demodulating the signal at the oscillation frequency of the tip agree well with calculations of the EM-LDOS evaluated at a height of around one micrometer, as shown in Fig. 15.16. This shows that, by demodulating the signal at Ω, those parts of the tip located relatively far from its apex contribute most of the TRSTM signal. The existence of contributions from regions situated some distance from the surface has been confirmed by producing approach–retract curves for the tip over distances of a few micrometers. In these experiments, the slow and gradual decrease of the TRSTM signal demodulated at Ω over a few micrometers has been observed [46].

As mentioned above, the TRSTM images of the EM-LDOS at a given energy, shown in Figs. 15.15b and 15.16, reveal the presence of fringes on the gold. These fringes arise due to electromagnetic surface waves called surface plasmons, which are thermally excited on the gold surface when the temperature is nonzero. These surface plasmons propagate like waves parallel to the surface of the material. The expression for the attenuation length of a plasmon propagating at the surface of a metal with dielectric constant ε_{m} at the interface with the air ($\varepsilon = 1$) is [51]

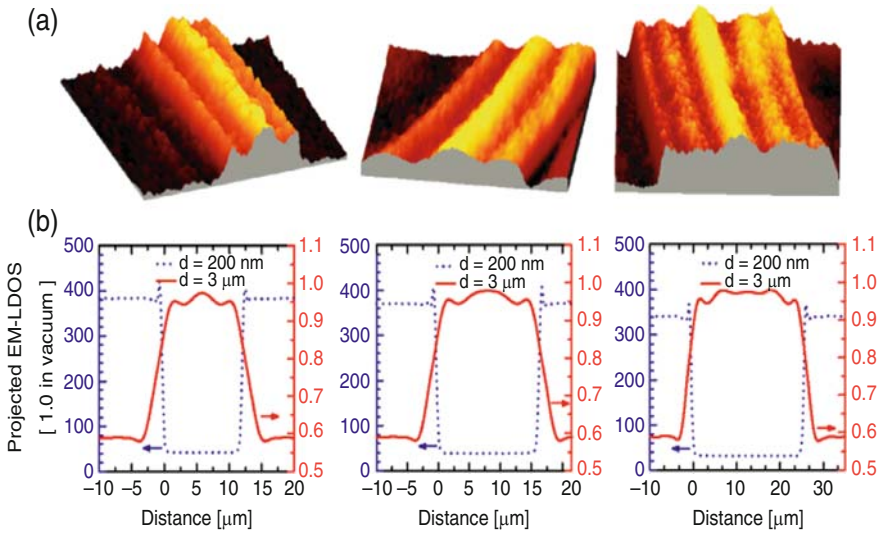

Fig. 15.16 Comparing (**a**) TRSTM measurements over different sections of the gold stripe with (**b**) theoretical calculations of the EM-LDOS projected along the tip axis and calculated at two different heights. In these experiments, the TRSTM signal is demodulated at the fundamental oscillation frequency Ω of the tip. Taken from [46]

$$L^* = \frac{1}{2\,\mathrm{Im}\,(\beta)}\,, \qquad \beta = k_0\sqrt{\frac{\varepsilon_\mathrm{m}}{\varepsilon_\mathrm{m}+1}}\,, \quad k_0 = \frac{\omega}{c} = \frac{2\pi}{\lambda}\,. \tag{15.18}$$

In the infrared range of our measurements, the surface plasmons on gold ($\varepsilon_\mathrm{m} = -3\,400 + 1\,300\mathrm{i}$ at $\lambda = 10$ μm [52]) have a propagation length of more than 1 mm. The confinement of surface plasmons thermally excited in the cavity defined by the geometry of the structure, which has a width of only a few tens of micrometers, produces interference patterns in the images of the EM-LDOS, analogous to those revealed by STM measurements of the electronic LDOS inside an atomic corral [53]. These results constitute the first direct proof of the partial spatial coherence of the near-field thermal radiation associated with surface waves. The phenomenon had been predicted and observed indirectly for the first time by J.-J. Greffet and coworkers by diffracting partially spatially coherent surface waves into the far field using a grating etched onto the sample [54].

Furthermore, when there are significant nonlinearities in the electromagnetic field close to the surface, the vertical extension of the tip region effectively contributing to the TRSTM signal can be reduced by demodulating the signal output by the infrared detector at a harmonic of the tip oscillation frequency, viz., 2Ω, 3Ω, etc. By applying this method, the vertical extension of the effective scatterer goes from a few micrometers when the signal is demodulated at Ω to about 200 nm when the signal is demodulated at 2Ω, as confirmed by the approach–retract curves of the tip.

An example of an image recorded on a sample comprising a gold stripe on SiC by demodulating the TRSTM signal at 2Ω is shown in Fig. 15.17. The TRSTM images

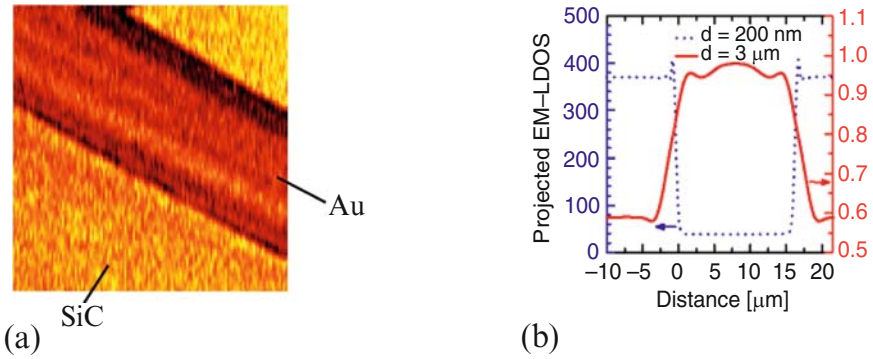

Fig. 15.17 (**a**) TRSTM image demodulated at 2Ω, obtained on a gold stripe deposited on an SiC substrate. (**b**) Theoretical calculation of the EM-LDOS for the same width of stripe. The EM-LDOS calculated at 200 nm is greater on the SiC than on the gold, in good agreement with experimental observations

demodulated at 2Ω exhibit contrast inversion compared with the TRSTM images demodulated at Ω. The TRSTM signal demodulated at 2Ω is stronger on the SiC than on the gold, in agreement with calculations of the EM-LDOS a few hundred nanometers from the surface.

This contrast inversion can be understood by looking at the dispersion curve for the surface phonon polaritons of the SiC, as shown in Fig. 15.18. Close to the frequency $\omega_{\max} = 178.7 \times 10^{12}$ s^{-1}, defined by the condition $\varepsilon'_{\mathrm{SiC}}(\omega_{\max}) = -1$ for a surface resonance to exist, there are a great many electromagnetic surface modes with different wave numbers but very close optical frequencies [48]. This behaviour

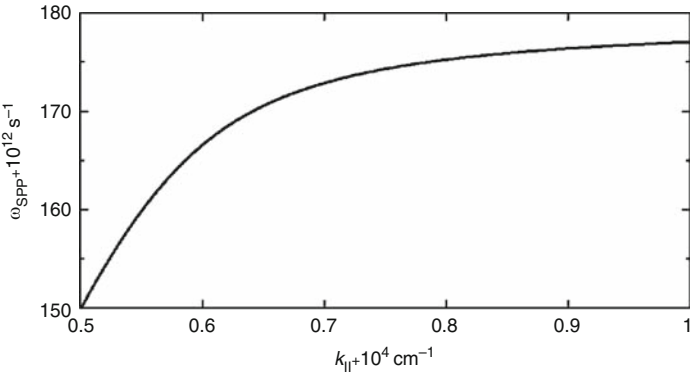

Fig. 15.18 Dispersion curve for surface waves on a plane interface between SiC and the vacuum. The horizontal asymptote gives rise to a resonance peak in the EM-LDOS and hence also in the electromagnetic energy density. For comparison, the spectral range of TRSTM measurements lies between 165×10^{12} s^{-1} ($\lambda = 11.4$ μm) and 181×10^{12} s^{-1} ($\lambda = 10.4$ μm). Taken from [48]

of the dispersion curve leads to the appearance of a quasi-monochromatic peak in the EM-LDOS at $\omega_{\max}$. The region containing $\omega_{\max}$ happens to coincide with the spectral range of TRSTM measurements. Theoretical calculations [48] show that the resonance at $\omega_{\max}$ in SiC produces a significant increase in the electromagnetic energy density at the surface, confined to just a few hundred nanometers in the direction normal to the surface. The contrast inversion between the SiC substrate and the gold stripe does not arise in TRSTM images demodulated at 2Ω when they are obtained on a substrate that does not have a resonance near 10.9 μm, or when a filter situated outside the SiC resonance is placed on the detector [46].

15.4.4 Prospects

Temperature Probe

Up to now we have shown that the TRSTM signal provides a measurement of the EM-LDOS at the tip position, and that the instrument has a spatial resolution of the order of 100 nanometers. The electromagnetic modes are associated with energy levels that are populated according to photon statistics, i.e., Bose–Einstein statistics. The temperature dependence of the TRSTM signal is perfectly known. It satisfies the relation in (15.17). In contrast with other local temperature probes, the TRSTM does not require contact with the surface, since it detects near-field thermal radiation from a material at a given temperature scattered by a dipole [46, 47].

Provided one knows the intensity $S(T_1)$ of the TRSTM signal measured at a given point, wavelength, and temperature T_1, e.g., $T_1 = 300$ K and $\lambda = 10$ μm, it should be possible to determine the temperature T_2 at this same point by measuring the TRSTM signal $S(T_2)$ at this temperature and at the same λ by referring to the curve of

$$\frac{S(T_2)}{S(T_1)} = \frac{\mathrm{e}^{h\omega/kT_1} - 1}{\mathrm{e}^{h\omega/kT_2} - 1}\,, \tag{15.19}$$

shown in Fig. 15.19. The TRSTM signal should thus allow an absolute determination of the local temperature at the sample surface.

Fourier Transform Spectroscopy with 100 nm Resolution

When designing the TRSTM, it was decided to avoid amplification of the signal by interference with a reference beam, and to concentrate on optimising photon collection and the sensitivity of the optical detection. The TRSTM does not then require a laser source and has a sensitivity exceeding that of an s-SNOM by several orders of magnitude.

TRSTM images of near-field thermal radiation recorded in the infrared ($\lambda \sim 10$ μm) on silicon carbide (SiC) or SiO_2 samples carrying gold (Au) patterns were obtained with a resolution that was sometimes as good as 100 nanometers [46]. The

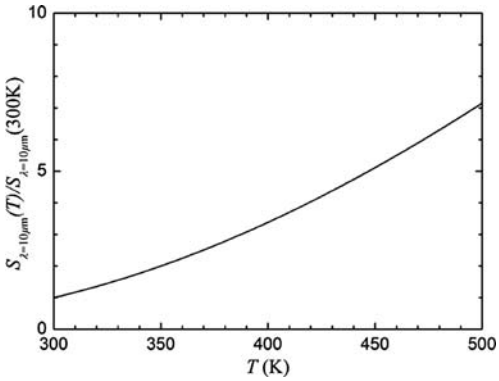

Fig. 15.19 Temperature dependence of the TRSTM signal $S_{\lambda=10\,\mu m}$ at $\lambda = 10$ μm, calculated assuming Bose–Einstein statistics

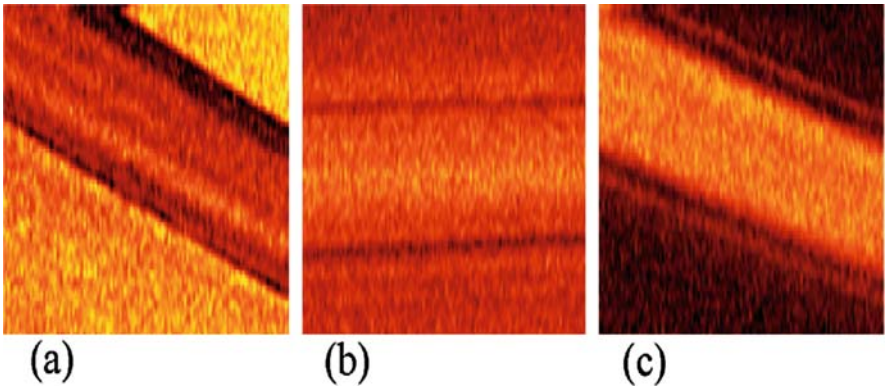

Fig. 15.20 TRSTM images recorded at the demodulation frequency 2Ω in order to reduce the vertical extension of the effective scatterer constituted by the tip. (**a**) Au stripe on an SiC substrate at $\lambda = 10.9$ μm (width of filter $\Delta\lambda = 1$ μm). (**b**) Au stripe on SiC at $\lambda = 8$ μm (width of filter $\Delta\lambda = 1$ μm). (**c**) Au stripe on SiO_2 at $\lambda = 10.9$ μm (width of filter $\Delta\lambda = 1$ μm). Taken from [46]

results shown in Fig. 15.20 were obtained by filtering the TRSTM signal using interference filters centered on different wavelengths and demodulating the signal at 2Ω. They show that the near-field thermal emission spectrum on SiC exhibits large variations depending on the wavelength, as predicted by the theory. It is theoretically predicted that the presence of surface waves such as plasmon polaritons or phonon polaritons should significantly alter the near-field thermal emission spectrum from the prediction of Planck's law, producing a quasi-monochromatic peak [48]. Comparing the images obtained on SiC (Figs. 15.20a and b) and on SiO_2 (Fig. 15.20c), it is clear that the spectrum also depends on the material under the tip.

A scanning probe like the 'electronic' STM is not only useful for its ability to make high resolution images, but also by the fact that it can carry out local spectroscopic measurements by varying the voltage between the tip and sample.

Nonlinearities in the characteristic of the tunnel current as a function of the voltage reveal invaluable information about the kind of material under the tip, or the existence of confined states.

Since the thermal radiation provides an internal source with extended spectrum produced by the sample itself, TRSTM is particularly useful for carrying out spectroscopic studies. There is thus no need for an external source with a broad spectrum, which constitutes the main technological bottleneck for carrying out spectroscopic studies using an s-SNOM in the infrared. As in 'electronic' STM, analysis of the near-field thermal emission spectrum detected with the TRSTM should provide a way of probing the energy dependence of the EM-LDOS locally.

Theoretical work [47] has shown that the near-field thermal radiation spectrum is given by

$$\rho(\omega) = \frac{1}{16\pi^2\omega z^3} \frac{\mathrm{Im}\left[\varepsilon(\omega)\right]}{\left|1+\varepsilon(\omega)\right|^2}\,, \tag{15.20}$$

which is none other than the electromagnetic local density of states (EM-LDOS). The Kramers–Kronig relations relate the real and imaginary parts of the dielectric function $\varepsilon(\omega)$, which completely characterises a material. From a near-field thermal radiation spectrum, it should be possible to invert the expression for the EM-LDOS to recover the behaviour of $\varepsilon(\omega)$. The inversion procedure has been validated theoretically for glass [55]. So by measuring the spectrum of the field scattered by the TRSTM tip, it should be possible to introduce a new kind of solid state spectroscopy operating in the mid-infrared and providing spatial resolutions in the 100 nanometer range.

Acknowledgements The authors are grateful for fruitful collaboration with J.-J. Greffet (Institut d'optique Graduate School, formerly at EM2C Ecole Centrale Paris), K. Joulain (LET, Futuroscope), R. Carminati (Institut Langevin, ESPCI), R. Colombelli (Inst. Electronique Fondamentale), V. Moreau (IEF), and A. Bousseksou (IEF). This work was supported by the French National Research Agency (ANR NanoFTIR) and the Centre de Compétence NanoSciences Ile de-France (project PSTS).

References

1. D.W. Pohl, W. Denk, M. Lanz: Appl. Phys. Lett. **44**, 651–653 (1984)
2. G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel: Phys. Rev. Lett. **50**, 120–123 (1983)
3. G. Binnig, C. Quate, G. Gerber: Phys. Rev. Lett. **56**, 930–933 (1986)
4. R. Bachelot, P. Gleyzes, A.C. Boccara: Microscopy Microanalysis Microstructures **5**, 389 (1994)
5. B. Knoll, F. Keilmann, A. Kramer, R. Guckenberger: Appl. Phys. Lett. **70**, 2667–2669 (1997)
6. H.-T. Chen, R. Kersting, G.C. Cho: Appl. Phys. Lett. **83**, 309–311 (2003)
7. A. Lahrech, R. Bachelot, P. Gleyzes, A.-C. Boccara: Opt. Lett. **21**, 1315–1317 (1996)
8. E. Hecht: *Optics*, 4th edn., Pearson Education Inc (2002)
9. S. Ducourtieux: Microscopie optique en champ proche sans ouverture, Ph.D. Thesis, University of Paris VI (2001)
10. L. Novotny, B. Hecht: *Principles of Nano-Optics*, Cambridge University Press, Cambridge (2006) pp. 38–41, 143

11. D. Courjon, C. Bainier (Eds.): *Le champ proche optique. Théorie et applications*, Springer-Verlag France (2001)
12. E.H. Synge: Philos. Mag. **6**, 356–362 (1928)
13. D.R. Turner: Etch procedure for optical fibers, United States patent 4,469,554 (1984)
14. P. Hoffmann, B. Duboit, R.-P. Salathé: Ultramicroscopy **61**, 165–170 (1995)
15. E. Neher, B. Sakmann: J. Physiol. (Lond.) **258**, 705–729 (1976)
16. J.A. Veerman, A.M. Otter, L. Kuipers, N.F. van Hulst: Appl. Phys. Lett. **72**, 3115–3117 (1998)
17. K. Karrai, R.D. Grober: Appl. Phys. Lett. **66**, 1842–1844 (1995)
18. M.A. Unger et al.: Etched chalcogenide fibers for near-field infrared scanning microscopy, Rev. Sci. Instrum. **69**, 2988 (1998)
19. D. Vobornik et al.: Infrared near-field microscopy with the Vanderbilt free electron laser: Overview and perspectives, Infrared Physics Technology **45**, 409–416 (2004)
20. M. Platkov et al.: Development of tapered silver-halide fiber tips for a scanning near-field microscope operating in the middle infrared, Rev. Sci. Instrum. **77**, 126103 (2006)
21. H. Bethe: Phys. Rev. (2) **66**, 163–182 (1944)
22. C.J. Bouwkamp: Philips Res. Rep. **5**, 321–332 (1950)
23. M. Platkov et al.: A scanning near-field middle-infrared microscope for the study of objects submerged in water, Appl. Phys. Lett. **92**, 104104 (2008)
24. B. Knoll, F. Keilmann: Optics Communications **182**, 321 (2000)
25. C.F. Bohren, D.R. Huffman: *Absorption and Scattering of Light by Small Particles*, John Wiley, New York (1983)
26. R. Bachelot et al.: Appl. Phys. Lett. **73**, 3333 (1998)
27. O.J.F. Martin et al.: Appl. Phys. Lett. **70**, 705 (1996)
28. L. Novotny et al.: Phys. Rev. Lett. **79**, 645 (1997)
29. T. Taubner et al.: Journal of Microscopy **210**, 311 (2003)
30. L. Gomez et al.: J. Opt. Soc. Am. B **23**, 823 (2006)
31. R. Hillenbrand et al.: Nature **418**, 159 (2002)
32. N. Ocelic et al.: Nature Materials **3**, 606 (2004)
33. T. Taubner et al.: Appl. Phys. Lett. **85**, 5064 (2004)
34. M. Brehm et al.: Nano Lett. **6**, 1307 (2006)
35. A. Lahrech et al.: Appl. Phys. Lett. **71**, 575 (1997)
36. A. Huber et al.: Adv. Mater. **19**, 2209 (2007)
37. T. Taubner et al.: Science **313**, 1595 (2006)
38. A. Huber et al.: Appl. Phys. Lett. **87**, 081103 (2005)
39. V. Moreau et al.: Appl. Phys. Lett. **90**, 201114 (2007)
40. N. Yu et al.: Appl. Phys. Lett. **91**, 173113 (2007)
41. A. Huber et al.: Nano Lett. **6**, 774 (2006)
42. M.M. Qazilbash et al.: Science **318**, 1750 (2007)
43. B. Cretin et al.: Scanning thermal microscopy. In: *Microscale and Nanoscale Heat Transfer*, ed. by S. Volz, Springer, Topics in Applied Physics Vol. 107 (2007) p. 181
44. L. Aigouy, Y. De Wilde, Ch. Fretigny: Les sondes thermiques locales. In: *Les Nouvelles microscopies. A la découverte du nanomonde*, Chap. 4, Editions Belin Echelles (2006)
45. L. Aigouy et al.: Appl. Phys. Lett. **87**, 184105 (2005)
46. Y. De Wilde, F. Formanek, R. Carminati, B. Gralak, P.-A. Lemoine, J.-P. Mulet, K. Joulain, Y. Chen, J.-J. Greffet: Nature **444**, 740 (2006)
47. K. Joulain, R. Carminati, J.-P. Mulet, and J.-J. Greffet: Phys. Rev. B **68**, 245405 (2003)
48. A. V. Shchegrov, K. Joulain, R. Carminati, and J.-J. Greffet: Phys. Rev. Lett. **85**, 1548 (2000); R. Carminati and J.-J. Greffet: Phys. Rev. Lett. **82**, 1660–1663 (1999)
49. G. Colas des Francs, C. Girard, J.-C. Weeber, C. Chicanne, T. David, A. Dereux, and D. Peyrade: Phys. Rev. Lett. **86**, 4950 (2001)
50. C. Chicanne, T. David, R. Quidant, J.C. Weeber, Y. Lacroute, E. Bourillot, A. Dereux, G. Colas des Francs, and C. Girard: Phys. Rev. Lett. **88**, 097402 (2002)
51. S.A. Maier: *Plasmonics. Fundamentals and Applications*, Springer (2007)
52. V.B. Svetovoy et al.: Phys. Rev. B **77**, 035439 (2008)

53. M.F. Crommie, C.P. Lutz, D.M. Eigler: Science **262**, 218 (1993)
54. J.-J. Greffet, R. Carminati, K. Joulain, J.-P. Mulet, S. Mainguy, and Y. Chen: Nature **416**, 61 (2002)
55. J.-J. Greffet et al.: *Coherent Spontaneous Emission of Light Due to Surface Waves*, Topics in Applied Physics Vol. 88, Springer (2003) p. 163

Chapter 16
PhotoThermal Induced Resonance. Application to Infrared Spectromicroscopy

Alexandre Dazzi

In this chapter, we shall describe a new technique for carrying out infrared spectromicroscopy at the nanoscale. This technique is based on the detection of expansion, induced by the photothermal effect, by the cantilever of an atomic force microscope (AFM). We begin by explaining the advantages of studying matter in the infrared frequency range, and the limits of conventional techniques. We then describe the underlying principles and the setup we have developed for going beyond these limits. We present a simple theoretical approach to the relevant phenomena in order to bring out the intrinsic properties of this method, and we discuss in particular the question of resolution. To end our discussion, we illustrate the potential of the PTIR technique by some experimental results.

16.1 Infrared Spectroscopy and Microscopy

In general, spectroscopy studies the way a wave interacts with matter, mainly through excitation and de-excitation phenomena, as a function of its energy or its frequency. In the infrared frequency range, information can be obtained about the molecular vibrations of matter by studying the absorption of electromagnetic waves. The frequency range 4 000–500 cm^{-1} (2.5–20 μm), referred to as the mid-infrared, is particularly interesting, because it contains all the vibrational bands of organic compounds. Infrared spectroscopy is thus an efficient tool for identifying the chemical composition of matter.

16.1.1 Optical Index and Absorption

In an infinite and transparent isotropic medium, an electromagnetic wave travels at a different speed v from the speed of light in vacuum, namely

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_16, © Springer-Verlag Berlin Heidelberg 2009

$$v = \frac{c}{n} , \tag{16.1}$$

where c is the speed of light in vacuum and n is the refractive index of the medium.

For an electromagnetic wave, the index of a medium can be written in the form

$$n = \sqrt{\mu_r \varepsilon_r} , \tag{16.2}$$

where μ_r is the relative magnetic permeability, taken equal to 1 in perfect dielectrics, and ε_r is the relative dielectric constant of the medium. When a medium is absorbent, it is convenient to introduce an imaginary part into the index. The optical index then takes the form

$$\tilde{n} = n + \mathrm{i}\kappa , \tag{16.3}$$

where n is the real refractive index and κ the extinction coefficient.

The intensity of an electromagnetic wave in an absorbing medium of thickness z, when reflection at the interfaces is neglected (the Beer–Lambert law), can then be written

$$I = I_0 \exp\left(-\frac{4\pi}{\lambda_0}\kappa z\right) , \tag{16.4}$$

where I_0 is the incident intensity, λ_0 the wavelength in vacuum, κ the extinction coefficient, and z the thickness of the medium. The intensity absorbed by the medium is therefore

$$I_{\mathrm{abs}} = I_0 \left[1 - \exp\left(-\frac{4\pi}{\lambda_0}\kappa z\right)\right] . \tag{16.5}$$

The transmittance T is defined as the ratio of the transmitted to the incident intensity, viz.,

$$T = \frac{I}{I_0} , \tag{16.6}$$

and the absorbance A is defined as

$$A = \log_{10}\frac{1}{T} . \tag{16.7}$$

Experimentally, what is observed spectrally is not a fixed value of the absorption, but rather a continuous variation called the absorption band. Indeed, molecular bonds can be simply modelled by mechanical models in which masses are joined together by springs. So when an electromagnetic wave excites the dipole field of the bond, it acts like an external oscillating force. The bond can thus be treated as a damped harmonic oscillator, forced at the frequency ω of the electromagnetic wave. In this case, the dependence of the extinction coefficient κ, associated with absorption, on the wave number σ, has the Lorentz profile

$$\kappa(\sigma) = \frac{\kappa_{\max}\Delta^2}{4(\sigma_0 - \sigma)^2 + \Delta^2} , \tag{16.8}$$

where $\kappa_{\max}$ is the maximum value of the coefficient at the resonance ($\sigma = \sigma_0$) and Δ the full-width at half-maximum of the band.

The resonance frequency ν_0 (wave number σ_0) is directly related to the type of molecular vibration, i.e., to the mass of the atoms taking part in the vibration and to the 'stiffness' of the bond between these atoms. The mechanical expression relating the resonance, mass, and stiffness is

$$\sigma_0 = \frac{1}{2\pi}\sqrt{\frac{f}{m}}\,, \tag{16.9}$$

where, in the present case, f is the stiffness of the molecular bond and m its reduced mass.

Finally, by analysing the absorption as a function of the frequency or the wave number, one discovers a certain number of bands characterising the molecular vibrations, and these bands will serve as a signature for the precise chemical composition of the given medium. By virtue of such results, one can thus detect or identify chemical compounds, or by relative analysis of the bands, determine the proportions of the constituents in the case of heterogeneous media.

16.1.2 Infrared Spectrometers

The optical instruments used to measure infrared spectra are spectrometers. They generally comprise a light source, a system for selecting the light frequency, and a detector. For many years, this kind of equipment used prisms to disperse the light. However, they consume a great deal of energy, and were replaced by gratings, which are much more efficient in terms of transmission. Today, most spectrometers are built using Fourier transform technology (see Fig. 16.1) [1–3]. The Fourier transform spectrometer (FTIR) is based on the Michelson interferometer. The light source is aimed at a beam splitter which sends 50% of the light to a fixed mirror M1 and the rest to the mirror M2 which can move along the axis. The two beams reflected by the mirrors converge on the sample, passing through the beam splitter, before reaching the detector.

For a monochromatic ray $T(\lambda)$, the intensity recorded by the detector will be

$$I = T(\lambda)\cos\varphi\,, \tag{16.10}$$

where $\varphi = 2\pi\delta/\lambda$, λ is the wavelength, and δ is the path difference between mirrors M1 and M2. The intensity of a monochromatic ray is thus a sinusoidal function of the path difference. It can also be expressed in the form

$$I(\delta) = T(\sigma)\cos(2\pi\sigma\delta)\,, \tag{16.11}$$

where σ is the wave number.

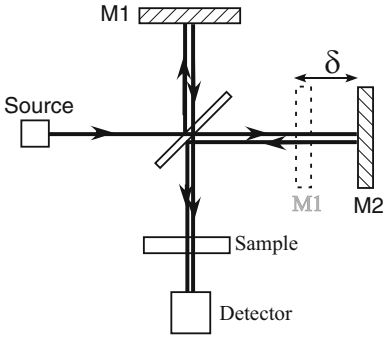

Fig. 16.1 Principle of the Fourier transform spectrometer

In order to describe an absorption band, rather than just a ray, the continuous dependence of the spectrum can be represented by an integral of $T(\sigma)$, whence the intensity becomes

$$I(\delta) = \int T(\sigma)\cos(2\pi\sigma\delta)\mathrm{d}\sigma\,. \tag{16.12}$$

The transmittance $T(\sigma)$ is found by carrying out the inverse Fourier transform:

$$T(\sigma) = \int I(\delta)\cos(2\pi\sigma\delta)\mathrm{d}\delta\,. \tag{16.13}$$

This shows the power of the Fourier transform arrangement. With a simply polychromatic source, and using the equivalence of the δ and σ spaces of the interferometer, one can produce a transmission spectrum. The spectrometer thus contains a computer which calculates the Fourier transform of the signal $I(\delta)$ to reconstruct the spectrum $T(\sigma)$.

The main advantage of this technology is that the measurement can be made very quickly, just by shifting the mirror in the interferometer. Another advantage is that the spectral resolution is constant over the frequency range, being related only to the displacements of the moving mirror. With current laser telemetry methods, the accuracy with which these displacements can be measured, and their high level of reproducibility, mean that spectral resolutions well below the cm^{-1} can be achieved. Furthermore, by repeating the measurements and averaging the signal, the signal-to-noise ratio can be considerably reduced, and by modulating the interferometer mirror, the linearity of the absorbance measurements remains valid over several decades. Finally, the digitisation of the spectra means that they can be easily manipulated and processed in real time (baseline correction, etc.).

To illustrate the way infrared spectral analysis is used, Fig. 16.2 shows the spectra of a bacterium and a polymer (PMMA) film. Note that the bands are broad for these two examples (a few tens of cm^{-1}), either because they comprise several types of vibration (stretching, deformation, or folding) of the same bond, or because they

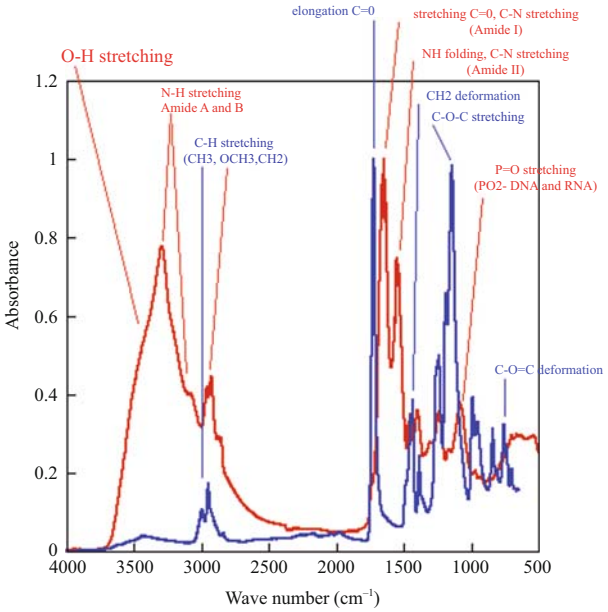

Fig. 16.2 Examples of infrared spectra obtained by FTIR spectroscopy of bacteria (*red*) and a poly(methyl methacrylate) (PMMA) film (*blue*). The main absorption bands are identified by their vibration modes

comprise several types of bond with neighbouring resonance frequencies. A biological entity such as a bacterium is much more heterogeneous chemically speaking than a polymer. This is reflected by the presence of extremely broad absorption bands, like the band at 3 300 cm^{-1}, which contains the vibrations of several bonds, viz., N–H, C–H, and O–H. It is nevertheless quite easy to distinguish these two objects by their very different infrared spectral signatures.

16.1.3 Confocal Microscopes

In contrast to the conventional microscope, the confocal microscope (see Fig. 16.3) is designed to produce an illumination focused on the sample and to image through an aperture. In this way, only those rays coming from the focal spot will be detected, and this provides better axial resolution. However, since the observation is made only at one point of the sample, the focal spot must scan over the whole surface in order to reconstitute the image. The advantage is that the focusing of the beam is used to reduce the depth of field and hence also the volume probed. The axial resolution is thus greatly improved over the conventional microscope. In addition, the signal-to-noise ratio of the images is improved by limiting the throughput of the beam.

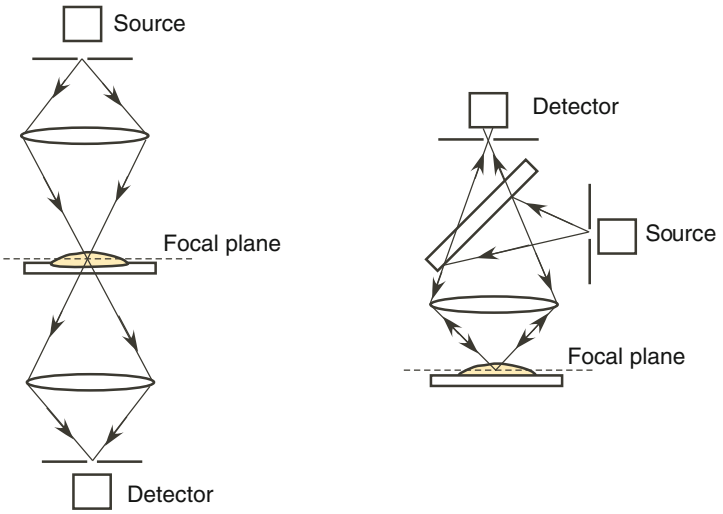

Fig. 16.3 Transmission (*left*) and reflection (*right*) confocal microscope

There are two configurations of the confocal microscope:

- The transmission configuration, which uses two focusing lenses.
- The reflection configuration, which uses a beam splitter, whence only one lens is needed.

Which arrangement is chosen often depends on the kind of sample to be studied, and the way it absorbs light. Indeed, when the transmission configuration is used, the sample must not be too thick to stop the light being transmitted into the detector.

But the main interest of this kind of setup is the fact that one can incorporate an infrared source upstream, coupled with a Fourier transform spectrometer. This coupling produces an arrangement known as a spectromicroscope, able to make absorption images for different wavelengths. Quite generally, for each sample, one can obtain a 3D acquisition volume described by two space coordinates (x,y) and a coordinate corresponding to the infrared spectrum. From such data, one could then study the spatial distribution of certain chemical compounds by fixing the infrared frequency on the corresponding absorption band and observing the resulting image. Alternatively, one could take cross-sections of the sample, by fixing some line in the (x,y) plane, and study the way the spectrum varies, i.e., the way the absorption bands appear and disappear as one moves along this line in space.

Regarding the spatial resolution, like any optical device, the confocal microscope is limited by diffraction. The size of the image of the hole in the diaphragm remains of the same order as the wavelength. For the infrared wavelengths that interest us, this means that the spatial resolution ranges between 2 and 20 μm. Furthermore, the light intensity getting through the diaphragm must be sufficient. This would not be the case with conventional light sources (black body). Only the synchrotron radiation to be found around large storage rings has allowed the diffraction limit to be

attained [4–6]. Recently, using semiconductor detector arrays (cooled to 77 K), allowing reasonable integration times, with conventional sources, it has been possible to achieve the same performance as with a synchrotron source [7, 8].

16.2 The PTIR Technique

The technique known as photothermal induced resonance (or PTIR) was developed to go beyond the characteristics of conventional spectromicroscopes and reach resolutions in the nanometer range. A patent (US 11/803421) has been deposited by the French research agency (CNRS), the University of Paris-Sud, and an American company. Conceptually, it seemed natural to replace the diffraction-limited confocal microscope by an optical near-field microscope, so as to gain access to sub-wavelength resolutions. A great deal of work [9–13] has shown that infrared absorption images and spectra can be obtained with different near-field microscope setups. The most remarkable work in this field has been done by Keilmann et al., resulting in the absorption image of a tobacco virus [14].

The main drawback with these optical devices is that they are difficult to implement. Furthermore, in the optical near field, the absorption signal is not so easy to extract and measured contrasts are not necessarily due to variations in the imaginary index. Indeed, light intensity contrasts can be caused by variations in the real index [15], which are related not only to the local characteristics of the sample, but also to surface height variations, i.e., the sample topography. The advantage with our photothermal method is that it is only sensitive to local absorption by the sample, more or less independently of its topography and density variations.

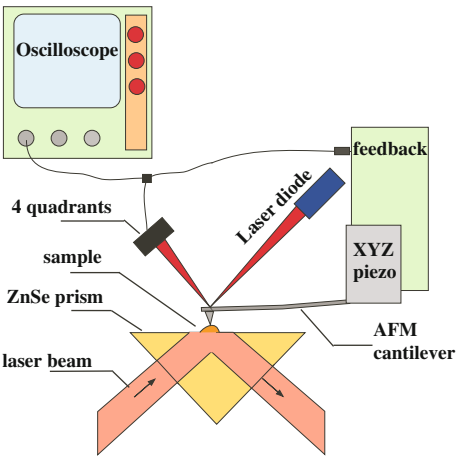

Fig. 16.4 Photothermal induced resonance (PTIR) setup

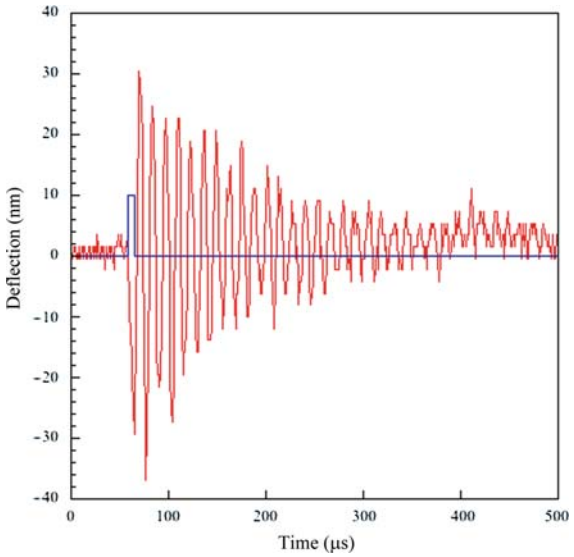

Fig. 16.5 Time dependence of the deflection of an AFM cantilever (*red curve*) for an 8 μs laser pulse (*blue curve*)

It was in this context that the PTIR technique was developed. The idea is to illuminate an object with a source of monochromatic light of the same frequency as one of the absorption bands of the object. If the wave is indeed absorbed, then the energy transmitted for excitation of this absorption band will be converted into a local temperature increase. This is the photothermal effect. If the temperature increases, the object will expand and generate a deformation field. This deformation will then be detected by a resonant system placed in direct contact with the object. We chose to use the cantilever of an atomic force microscope (AFM), whose tip is held in contact with the object (see Fig. 16.4). For each deformation of the object, the cantilever will thus begin to oscillate at its resonance frequencies. The oscillations are then detected by the conventional AFM optical arrangement (laser diode and four quadrants) and recorded on an oscilloscope.

Figure 16.5 shows the typical oscillation of the AFM cantilever when it is in contact with an absorbing object following a laser pulse. Note that the oscillation period is of the order of 20 μs, which is much longer than the laser pulse (< 10 μs). The damping of the oscillations, generally extending over times of the order of a few hundred microseconds, is directly related to the way the tip rubs against the observed object. By analysing the cantilever oscillations, it is thus possible to deduce the local absorption by the sample. Hence this method can be used to measure either the absorption spectrum when the wavelength is changed for a fixed measurement point, or absorption maps by doing the opposite. As in confocal microscopy, this is achieved by scanning the tip over the whole surface (the illuminated area being much bigger than the region scanned by the AFM).

In the rest of this chapter, we shall discuss how the measured signal (cantilever oscillations) can be related to the absorption by the sample, using a simple description of the relevant physical phenomena.

16.3 Photothermoelastic Phenomena

In order for our system to function correctly, pulsed infrared lasers must be used so that the deformation induced by the photothermal effect is as fast as possible. Indeed, the temperature variation must be faster than the reaction time of the AFM feedback (ms). When an object absorbs infrared electromagnetic radiation, it can be assumed that all the absorbed energy is converted into temperature by the material (the vibrational excitation of the bond is damped by collisions with the other atoms situated in the vicinity of the center of absorption). The temperature increase is therefore

$$\Delta T = \frac{E_{\text{abs}}}{\rho V C_{\text{p}}} , \tag{16.14}$$

where C_{p} is the specific heat capacity of the material, ρ is the density, V is the volume, and E_{abs} is the absorbed energy.

For a 1D object of length x, the extension Δx is then given as a function of the temperature rise by

$$\frac{\Delta x}{x} = \alpha \Delta T = \frac{\alpha}{\rho V C_{\text{p}}} E_{\text{abs}} , \tag{16.15}$$

where α is the thermoelastic coefficient. For example, for an object of radius 1 μm and a temperature rise ΔT of 10 K, the vertical displacement is 1 nm (taking the thermoelastic coefficient to be equal to 10^{-4} K^{-1}). This vertical displacement is perfectly detectable by an AFM (vertical sensitivity of a few Å).

16.4 Time Scales

Quite generally, the deformation dynamics of absorbing objects will depend not only on the thermal and mechanical parameters of the various materials, but also on the duration of the laser pulse. All calculations have been done using the COMSOL software. They will help us to understand the relevant orders of magnitude and identify the key parameters determining these phenomena.

In the rest of the chapter, we shall be interested in two specific cases (see Fig. 16.6):

- Case 1. An absorbent object with square cross-section of side s, placed on a surface that is transparent to the infrared (ZnSe) and surrounded by air.

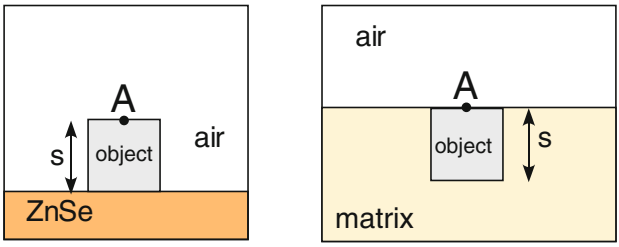

Fig. 16.6 Case 1. Absorbing square object (*grey*) of side *s* placed on a ZnSe surface that is transparent to infrared radiation. Case 2. Absorbing square object (*grey*) of side *s* buried in a matrix with the same thermomechanical properties. In both cases, the temperature and expansion calculations are carried out at point A, i.e., in the middle of the upper face of the square

- Case 2. An absorbent object with square cross-section of side *s*, buried in an infinite matrix with the same thermomechanical properties, in such a way that one face of the square is in contact with the air.

These are the two most representative cases for modelling our samples. Either the sample is isolated on a working surface, or the absorbing object is buried in a non-absorbing structure. In all calculations, we shall take the mechanical and thermal properties of poly(methyl methacrylate) (PMMA), a polymer that is fairly representative of organic materials.

To model the absorption, we shall assume that the laser induces a heat source Q over a lapse of time t_0 (box function), and this solely on the absorbing object (grey in Fig. 16.6). Then the temperature rise and deformation of the object can be calculated. We have chosen to represent only the deformation of the middle of the object (point A) to simplify the description.

Figure 16.7 shows the time dependence of the expansion Δx for three different laser pulses (with constant total energy input Qt_0) for case 1. In the three cases, it can be observed that there is already a rapid increase in expansion which lasts as long as the laser illuminates the absorbing object. Then the deformation returns to equilibrium in a time t_{relax} which is directly imposed by thermal diffusion. Note that the time t_{relax} is always much longer than the time t_0 indicated in the graph. In this case, thermal diffusion phenomena do not have time to enter into competition with the energy supplied by the source. This is why the object expands linearly during the illumination stage.

The total deformation time can thus be described as the sum of the duration of the laser pulse and the relaxation time of the object in the approximation $t_0 < t_{\text{relax}}$:

$$t_{\text{def}} = t_0 + t_{\text{relax}} \,. \tag{16.16}$$

It is easy to imagine that, if the illumination time t_0 is much longer than t_{relax}, one will reach an equilibrium situation in which the object will remain dilated until the

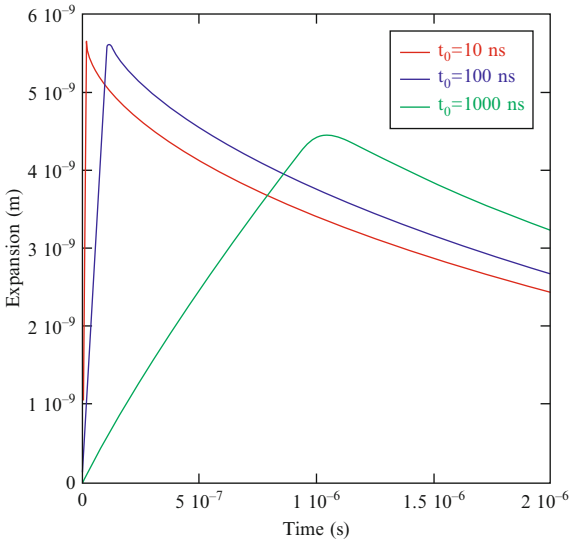

Fig. 16.7 Expansion profile at point A of the square object (case 1) as a function of time, for different illumination times t_0 by the infrared source

illumination is switched off. So if $t_{\text{relax}} \ll t_0$, the time dependence of the deformation can be taken as the time dependence of the laser pulse.

As just observed, it is therefore the relaxation time of the object that will impose the dynamics of the response to illumination. For a sphere of radius R, this relaxation time is given by

$$t_{\text{relax}} = \frac{\rho C_{\text{p}}}{3k_{\text{th}}} R^2 = \frac{R^2}{D} \,, \tag{16.17}$$

where k_{th} is the thermal conductivity of the sphere, ρ the sphere density, and D the equivalent thermal diffusivity.

In the case of an object made from a polymer like PMMA, the thermal diffusivity D is of order 10^{-7} m^2/s. When the time required for the exponential decrease of the deformation is calculated numerically in case 1 as a function of the size s of the absorbing object, a similar relation is found (see Fig. 16.8):

$$t_{\text{relax}} = \frac{s^{1.98}}{D_{\text{case 1}}} \,. \tag{16.18}$$

As the square object lies on a surface, the power here is not exactly equal to 2, but it is nevertheless very close. This means that the relaxation process is fairly similar to that of a sphere. Regarding the diffusivity $D_{\text{case 1}}$, it is equal to 4.2×10^{-7} m^2/s. It is thus slightly higher than for the sphere, because the square is in contact with a surface that has a thermal conductivity 100 times higher than PMMA.

These simple calculations show us that the relevant orders of magnitude are rather close to those for a sphere, and hence that this simple model can already provide us

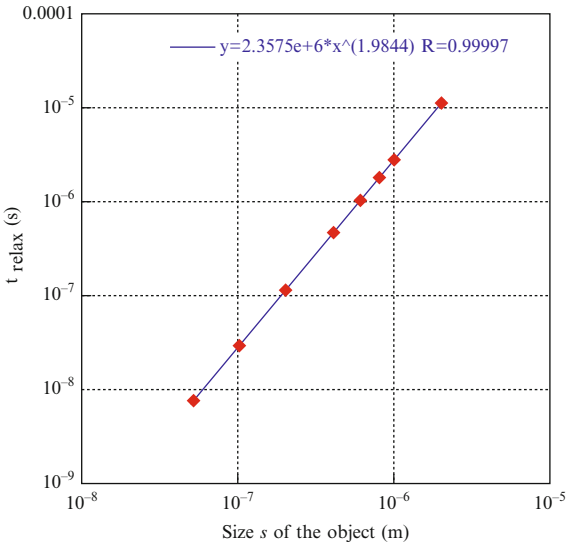

Fig. 16.8 Dependence of the relaxation time (required for return to equilibrium) on the size s of the square object in the situation of case 1. Calculation results are shown in *red* and the fitted power law curve in *blue*

with a good indication of the relevant time scales. For a square of side 1 μm, we find $t_{\text{relax}} = 2.43$ μs, whereas for a square of side 100 nm, $t_{\text{relax}} = 28.8$ ns. We thus see that the relaxation times can become extremely short, of picosecond order, for objects with size around the nanometer.

Regarding case 2, the relaxation of the object no longer follows a purely exponential decrease, but is better described by the sum of two decreasing exponentials. We thus define two relaxation times t_1 and t_2. Figure 16.9 shows that the relation between these characteristic times and the size of the object is still a power law:

$$t_1 = \frac{s^{1.91}}{D_1}\,, \qquad t_2 = \frac{s^{1.95}}{D_2}\,. \tag{16.19}$$

Insofar as the object is buried, the power is close to 1.9 and the diffusivity coefficients are $D_1 = 9.4 \times 10^{-7}$ and $D_2 = 3.2 \times 10^{-8}$ m^2/s, respectively. The time t_1 corresponds mainly to the relaxation of the absorbing object. We observe that it has better diffusivity than in case 1. This can be explained by the fact that it is surrounded by a matrix with the same thermal conductivity. It will thus be easier to evacuate heat into the matrix than into the air. The second characteristic time t_2 is associated with the way the matrix evacuates the heat released into it by the absorbing object. This time is also related to the size of the object, since it determines the amount of heat to be evacuated. We observe that the diffusivity is 30 times lower in this case. This can be explained by the fact that the infinite matrix will take longer to dilute this excess heat.

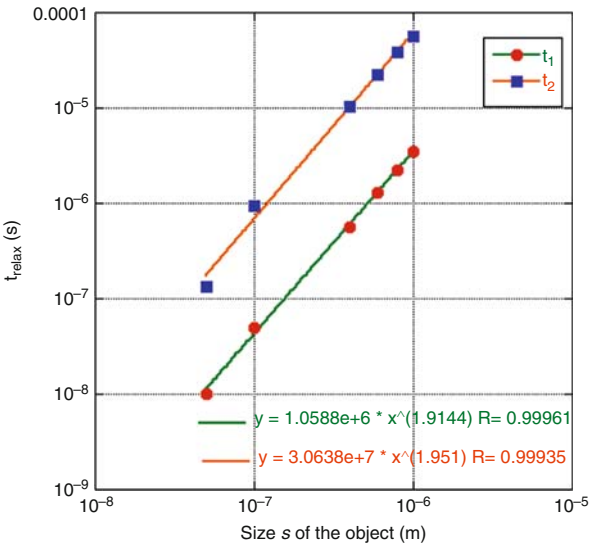

Fig. 16.9 Dependence of the relaxation time on the size s of the square object in the situation of case 2. *Red dots* correspond to short relaxation times and *blue squares* to long relaxation times. *Green* and *orange curves* are fitted power law curves

In both the presented cases, the orders of magnitude of the relaxation times for objects of micron size are around a few microseconds or a few tens of microseconds. The relaxation times for objects of size 100 nm are of the order of a few tens of nanoseconds or a few hundred nanoseconds.

16.5 Thermoelastic Deformation

In the 1D case, the expression for the expansion induced by a temperature rise is given by (16.15). Here we have checked that this linear law can also be applied to cases 1 and 2 by defining a coefficient C determined by the surroundings, which modifies the apparent value of the thermoelastic coefficient. Equation (16.15) becomes

$$\frac{u_0}{s} = \alpha C \Delta T \,, \tag{16.20}$$

where u_0 is the deformation at the center of the object (point A), s is the size of the object, and ΔT is the temperature rise.

In cases 1 and 2, the coefficients C are 1.528 and 1.543, respectively. These values are extremely close and show that the expansion of the square is barely modified by the presence or otherwise of stresses exerted on its edges. Provided that the upper surface remains free, the expansion is the same. Only the characteristic time scales involved are different (see the last section).

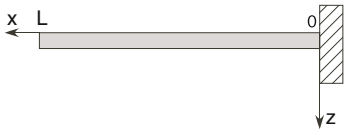

Fig. 16.10 Beam of length L, clamped at $x = 0$

16.6 AFM Contact Resonance Mode

To detect the expansion of the object, AFM is used in contact mode. Under the effect of the expansion, the cantilever is lifted up and begins to vibrate. To determine the possible vertical deformation modes of a cantilever of simple shape and clamped at one end, as shown in Fig. 16.10, we use Newton's equation

$$EI\frac{\partial^4 z}{\partial x^4} + \rho A\frac{\partial^2 z}{\partial t^2} = 0\,, \tag{16.21}$$

where E is Young's modulus, I is the quadratic moment of inertia of the beam, ρ is its density, A its cross-sectional area, and z the vertical deformation.

The solutions of such an equation have the form

$$z(x,t) = \left(a_1 \mathrm{e}^{\beta x} + a_2 \mathrm{e}^{-\beta x} + a_3 \mathrm{e}^{\mathrm{i}\beta x} + a_4 \mathrm{e}^{-\mathrm{i}\beta x}\right)\mathrm{e}^{\mathrm{i}\omega t} = g(x)\mathrm{e}^{\mathrm{i}\omega t}\,. \tag{16.22}$$

In the present case, the beam is fixed at $x = 0$. The contact between the AFM tip and the object is modelled by a spring of stiffness k_c placed at the end of the beam, i.e., at $x = L$. The boundary conditions are:

$$\begin{aligned} &g(x) = 0\,, \quad &&\frac{\partial g(x)}{\partial x} = 0\,, \quad &&\text{at } x = 0\,, \\ &\frac{\partial^2 g(x)}{\partial x^2} = 0\,, \quad &&\frac{\partial^3 g(x)}{\partial x^3} = \frac{k_\mathrm{c}}{EI} g(x)\,, \quad &&\text{at } x = L\,. \end{aligned} \tag{16.23}$$

At $x = 0$ (clamped end), the boundary conditions reflect the fact that the displacement is zero and that the resulting slope of the deformation is also zero. At $x = L$ (free end of the beam), it is assumed that the bending moment is zero and that the shear stress is proportional to the force acting on the free end of the beam. In the present case, the end of the AFM cantilever (the tip) is in contact with the surface. This contact can be modelled by an elastic bond behaving like a spring. The constant k_c then represents the contact stiffness.

Implementing the four equations, we obtain the eigenvalue equation for this system:

$$\cos(\beta L)\sinh(\beta L) - \sin(\beta L)\cosh(\beta L) = \frac{k_0(\beta L)^3}{3k_\mathrm{c}}\left[1 + \cos(\beta L)\cosh(\beta L)\right]\,, \tag{16.24}$$

Table 16.1 First seven solutions to the eigenvalue equation (16.24)

Solution	Value
$\beta_0 L$	3.9266
$\beta_1 L$	7.0686
$\beta_2 L$	10.2102
$\beta_3 L$	13.3518
$\beta_4 L$	16.4934
$\beta_5 L$	19.6350
$\beta_6 L$	22.7764

Table 16.2 First seven oscillation frequencies of a silicon AFM cantilever with characteristics $L = 450$ μm, $E = 150$ GPa, $b = 2$ μm, $r = 2330$ kg/m^3

Mode	Frequency [kHz]
0	56.13
1	181.91
2	379.55
3	649.05
4	990.43
5	1403.67
6	1888.74

where k_0 is the stiffness constant of the AFM cantilever and L is its length. For a perfectly bound tip ($k_c = \infty$), equation (16.24) can be simplified by noting that

$$\tan(\beta L) = \tanh(\beta L) \,, \tag{16.25}$$

The solutions thus form a discrete set of values that can be found numerically. The first 7 values are exhibited in Table 16.1.

The relation between the frequency and the mode number is obtained from (16.21):

$$\omega_n = \beta_n^2 \sqrt{\frac{EI}{\rho A}} = \beta_n^2 \sqrt{\frac{Eb^2}{12\rho}} \,, \tag{16.26}$$

where b is the thickness of the beam, E its Young's modulus, and ρ its density. To get some idea of the frequency values reached by these modes, Table 16.2 shows the results obtained by calculation for a silicon cantilever with the characteristics $L = 450$ μm, $E = 150$ GPa, $b = 2$ μm, $r = 2330$ kg/m^3.

The fundamental contact mode is of the order of about ten kHz. The period associated with these frequencies is 17.8 μs for mode 0 and 5.5 μs for mode 1. Note that these times are longer than the relaxation times of objects of micron size (see Sect. 16.4). This means that the surface deformation could not be described by oscillations of the cantilever. If we are to be able to detect the return to equilibrium of an object that relaxes in 1 μs, we require at least ten oscillations to obtain a good description of the decay, which means that we must have a cantilever with a contact

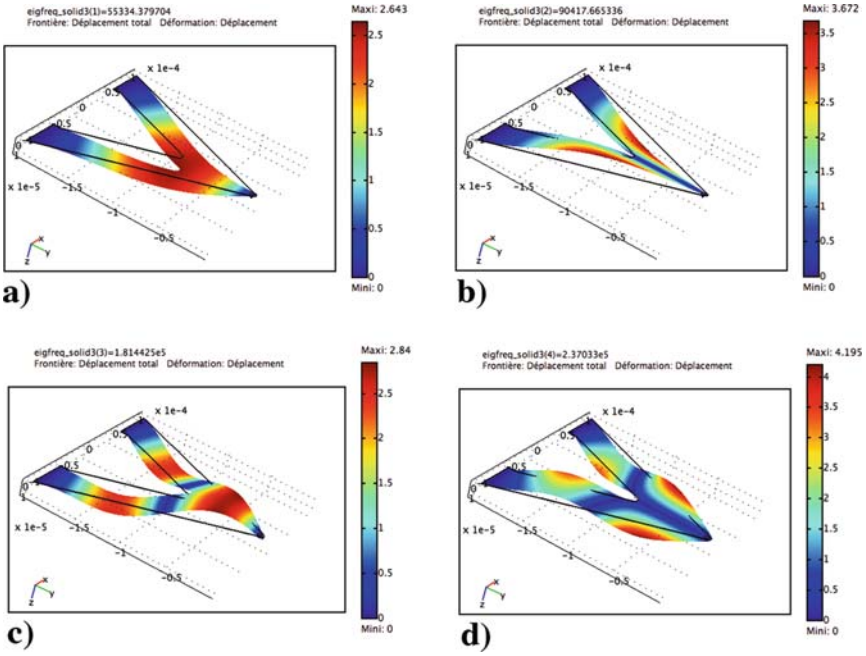

Fig. 16.11 Representation of the first 4 vibration modes for contact resonance. The colour code indicates the deformation field of each mode. *Blue* represents zero and *red* a maximal deformation

mode of 10 MHz. In our situation, it is better to use flexible cantilevers, hence with lower frequencies, so that the AFM is sensitive to small force variations. Then the perturbations generated by the expanding surface will be perceived by the AFM cantilever as instantaneous excitations (Dirac delta functions).

When we need to calculate the vibration modes of cantilevers with more complex geometrical shapes, such as the V-shaped silicon nitride (Si_3N_4) cantilevers, it is convenient to use a computation software implementing the finite-element method, which has the advantage that it can perfectly model the tip. Figure 16.11 shows the first 4 modes of a V-shaped cantilever with a stiffness constant of 0.1 N/m, with the boundary condition that the tip should not move in the vertical direction (the z direction), but rather in the (x,y) plane. The calculated values of the mode frequencies are 55.3 kHz, 90.4 kHz, 181.4 kHz, and 237 kHz. They are fairly close to the experimental values, viz., 60 kHz, 89 kHz, 180 kHz, and 210 kHz.

With this kind of approach, the influence of the tip contact conditions can also be modelled. For a better description of the tip motion at the surface, we impose boundary conditions on the loads in the three space directions at the end of the tip:

$$F_x = k_x u\,, \qquad F_y = k_y v\,, \qquad F_z = k_z w\,, \tag{16.27}$$

where k_x, k_y, and k_z are the contact stiffnesses in the three directions, and u, v, and w are the corresponding displacements. In this way, the contact is characterised by

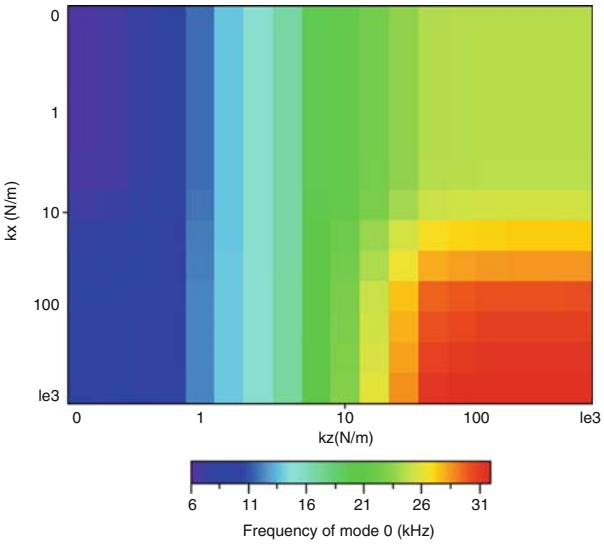

Fig. 16.12 Contact resonance frequency of the fundamental mode of a V-shaped cantilever of stiffness 0.03 N/m as a function of the vertical and horizontal contact stiffnesses, k_z and k_x, respectively

3 stiffness values. Generally, the stiffness in the (x,y) plane is isotropic, whence $k_x = k_y$. The influence of the contact stiffnesses shows up directly on the values of the contact resonance frequencies. This is already suggested by the expression in (16.24), where k_c, playing the role of k_z here, determines the eigenvalues of the modes.

Figure 16.12 shows the influence of the values of k_x and k_z on the value of the fundamental mode for a V-shaped cantilever with stiffness constant 0.03 N/m. When the values of k_x and k_z are zero, we recover the vibration frequency in air (6 kHz). When k_z is infinite and k_x zero, we find the contact frequency (24.7 kHz) associated with rigid contact, i.e., no displacement in the z direction, with perfect slipping, i.e., free displacement in the (x,y) plane. When k_x becomes infinite, the tip is perfectly immobilised on the surface, and the contact frequency is higher (32.3 kHz). Changing the value of k_x leads to rather significant shifts in the resonance, up to a shift of 7.6 kHz between perfect slipping in the (x,y) plane and immobilisation. This shows that the contact mode will be extremely sensitive to the way in which the tip rubs against the sample surface.

16.7 Absorption Measurement by Contact Resonance

Only the expansion (impulse) will generate energy in the cantilever. The relaxation time required for return to equilibrium only provides an indication of the maximal frequency of the excited normal modes. For example, if the relaxation time is of

microsecond order, then none of the contact modes greater than the MHz will be excited by the expansion. Considering characteristic deformation times $t_{\mathrm{def}} = t_0 + t_{\mathrm{relax}}$ of the absorbing objects that are much shorter than the period $T_0 = 2\pi/\omega_0$ of the contact eigenmodes of the AFM cantilever, the displacement speed of the surface of the object is

$$v_0 = \frac{u_0}{t_0} , \tag{16.28}$$

where u_0 is the thermoelastic deformation induced by the temperature rise (16.20) and t_0 is the pulse width of the laser.

The time dependence of the contact mode can be described by the differential equation of motion of a damped mass–spring system:

$$\frac{\partial^2 z}{\partial t^2} + \eta \frac{\partial z}{\partial t} + \omega_0^2 z = 0 , \tag{16.29}$$

where η is the damping coefficient, ω_0 the angular frequency of the contact mode, and z the deflection of the cantilever. Note that z does not here describe the displacement of the tip in the surface, but the oscillation of the cantilever arm above the tip (swinging flexure motion). Since the AFM is in contact, there is no vertical displacement.

The initial conditions of the cantilever motion are imposed by the displacement of the surface of the absorbent object, conditioned by the duration t_0 of the laser pulse:

$$\dot{z}(t=0) = \dot{z}_0 , \qquad z(t=0) = z_0 = \dot{z}_0 t_0 . \tag{16.30}$$

The solution of (16.29) taking into account the boundary conditions (16.30) is

$$z(t) = \left(\frac{\dot{z}_0}{\omega_0} + \frac{\eta}{2\omega_0} \dot{z}_0 t_0 \right) \mathrm{e}^{-\eta t/2} \sin(\omega_0 t) + \dot{z}_0 t_0 \mathrm{e}^{-\eta t/2} \cos(\omega_0 t) . \tag{16.31}$$

If the period of the contact mode is much longer than the duration of the laser pulse, i.e., $T_0 \gg t_0$, the expression in (16.31) simplifies to

$$z(t) = \frac{\dot{z}_0}{\omega_0} \mathrm{e}^{-\eta t/2} \sin(\omega_0 t) . \tag{16.32}$$

In fact, everything happens as though the initial displacement due to the thrust of the surface is negligible, and only the initial impulse is dominant (only sensitive to the impact).

If the damping can be assumed slow enough to allow the cantilever to oscillate several times, the first maximum of the oscillations can be approximated by

$$z_{\max} = \frac{\dot{z}_0}{\omega_0} = \frac{z_0}{\omega_0 t_0} . \tag{16.33}$$

The cantilever is sensitive to the applied force. This means that the initial displacement z_0 of the cantilever is related to the variation of the initial force acting on it through the expansion of the surface. Equation (16.33) can thus be written

$$z_{\max} = \frac{\dot{z}_0}{\omega_0} = \frac{z_0}{\omega_0 t_0} = \frac{F_0}{\omega_0 t_0 k_0} , \tag{16.34}$$

where k_0 is the stiffness of the cantilever.

If we assume that the contact area with the tip is defined by a contact radius a (imposed by the static force required to operate the contact mode), the initial force F_0 induced by the deformation can thus be written

$$F_0 = \pi a^2 \sigma_0 , \tag{16.35}$$

where σ_0 is the pressure induced by the displacement u_0. The relation between the displacement and the strain is

$$\sigma_0 = E\varepsilon = E\frac{u_0}{s} , \tag{16.36}$$

where ε is the relative extension and E is Young's modulus. Substituting (16.35) and (16.36) into (16.34), this yields

$$z_{\max} = \frac{\pi a^2 E}{\omega_0 t_0 k_0}\frac{u_0}{s} . \tag{16.37}$$

We thus find that the oscillation of the cantilever is proportional to the surface displacements of the absorbing object. Note also that the shorter the impulse (for total constant energy), the larger will be the amplitude of the oscillations. This property stems from the detection sensitivity of the contact modes for impulsive excitations.

When the surface deformation is produced by local heating, the displacement u_0 in (16.37) can be replaced by the temperature rise of the object in (16.20). The expression then becomes

$$z_{\max} = \frac{\pi a^2 E}{\omega_0 t_0 k_0}\alpha C \Delta T . \tag{16.38}$$

The maximal oscillation amplitude of the contact mode is once again perfectly proportional to the temperature rise. The linearity of the physical relations necessarily induces the linearity of the expressions. This tells us that temperature measurements by contact mode must remain within a range of small induced displacements and hence small temperature changes.

At the beginning of the chapter, we saw that the absorption of infrared photons would generate a temperature rise by the photothermal effect (16.14). Equation (16.38) can thus be written in terms of the absorbed light energy:

$$z_{\max}(\sigma) = \frac{\pi a^2 E}{\omega_0 t_0 k_0}\frac{\alpha C}{C_{\mathrm{p}}\rho V}E_{\mathrm{abs}}(\sigma) . \tag{16.39}$$

Each absorption band can be described as a change in the absorbed energy as a function of the frequency σ. By detecting oscillations in the contact mode, the absorption spectrum of the sample can thus be measured directly. The advantage with

this method is that it provides a way round the problems of the optical near field, and a way of obtaining a spectrum by direct measurement.

Detection sensitivity is therefore limited by the detection sensitivity of the AFM. As an example, consider the absorption by an object made from PMMA, with size 1 μm, placed on a ZnSe surface. A temperature rise of 6 K is imposed, corresponding to absorption of 1 nJ by the object. We set $a = 3$ nm, $E = 3$ GPa, $\omega_0 = 2\pi \times 50\,000$ Hz, $t_0 = 50$ ns, $k_0 = 0.1$ N/m. Under these conditions, the amplitude $z_{\max}$ is equal to 34 nm for a surface deformation of 0.65 nm. Note that this detection system amplifies the motion, since it increases here by a factor of about 50.

Free oscillations of the cantilever will be directly defined by the thermal noise at room temperature. The sensitivity limit will thus be related to these thermal oscillations at room temperature:

$$\left\langle z_{\text{th}}^2 \right\rangle = \frac{k_{\text{B}} T}{k_0} \,, \tag{16.40}$$

where k_{B} is Boltzmann's constant and T the temperature in kelvin. Applying this expression to our system, we find that the amplitude z_{th} of thermal oscillations is 0.4 nm. For example, returning to the case of the PMMA object above, this leads to a detection limit equal to a temperature rise of 70 mK!

In practice, we do not measure the amplitude of the first oscillation, but rather the amplitude of the Fourier transform of the deflection signal. Through this analysis, the noise in the signal is filtered, whence the sensitivity of the deformation detection can be improved to a few hundred picometers, corresponding to a few tens of mK.

The PTIR technique is thus found to be extremely sensitive to fast temperature increases, so it is not necessary to inject much energy into the absorbing object in order to detect an effect.

16.8 PTIR Lateral Resolution

As we saw in the last section, PTIR can measure the infrared absorption by an object through the oscillations of an AFM cantilever. By fixing the wavelength of the illumination on a specific absorption band, corresponding to a particular molecular vibration, the tip can be scanned across the surface of the object to probe the distribution of absorption at that wavelength. This type of analysis is called chemical mapping or cartography. As noted above, the expansion of the object is proportional to the absorption. So when we consider the values of the oscillation amplitude at different points of the surface, what we obtain is simply an image of the maximal deformation of the object, since the way it returns to equilibrium is too fast to be detected. All the expansion images obtained like this will be absorption images.

16.8.1 Resolution of an Object Placed on a Surface

This situation corresponds to case 1 of Sect. 16.4. As we have seen, the deformation of an object via the photothermal effect is directly related to its thermoelastic characteristics. For most organic materials, the thermoelastic coefficient lies in the range 10^{-4}–10^{-5} K^{-1}. This means that, for a temperature rise of 10 K, the object will deform by at most 0.1%. This relative deformation is really very small indeed, since it corresponds to an expansion of 1 nm for a size of 1 μm. We thus see that the increase in the size of the object is negligible compared with its size, so that the topographic image can be considered very close to its absorption image. Hence the expansion will not destroy the detectability of an object (point spread function), in contrast to conventional optical systems in which a point source appears with a width equal to the wavelength. Under these conditions, it is easy to understand why the ultimate resolution of chemical mapping images for objects deposited on surfaces will be the same as that obtained for topographic images. Since AFM systems with conventional tips have resolutions of 10 nm (for Si cantilevers) to 50 nm (for Si_3N_4 cantilevers), it is therefore quite possible to achieve this level of resolution, and this whatever the wavelength used.

16.8.2 Resolution of a Buried Object

This situation corresponds to case 2 of Sect. 16.4. When the object is buried, the expansion will not only push the free face on the air side, but also drag with it the region of the matrix surrounding it. The resulting deformation will thus appear larger than the actual size of the object (see Fig. 16.13). In order to determine and

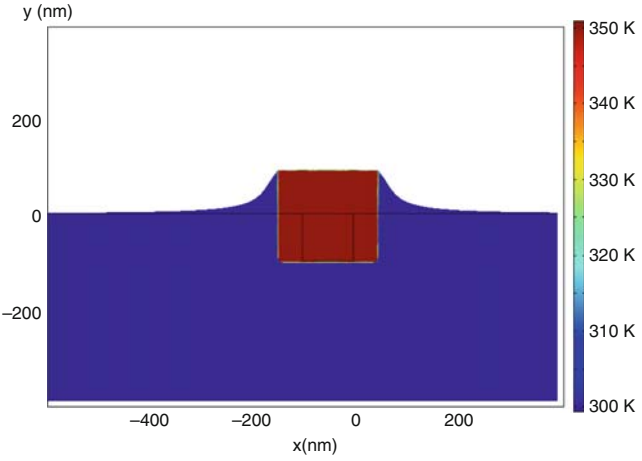

Fig. 16.13 Representation of the expansion of a square object of side 100 nm buried in a matrix (case 2). *Red* corresponds to a temperature of 350 K and *blue* to a temperature of 300 K. The representation of the deformation has been magnified by a factor of 2 to make it more visible

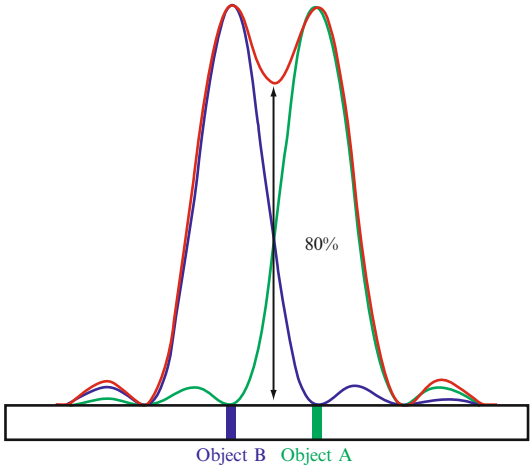

Fig. 16.14 Representation of the light intensity distribution of two point objects, A (*green*) and B (*blue*), together with the corresponding total distribution (*red*)

understand the role of the parameters governing the resolution, we may use the same criterion as in optics, viz., the Rayleigh criterion. The resolution is thus defined as the distance between two identical objects such that the thermomechanical deformation field is able to distinguish between them. In our examples, the geometry of the problem imposes translation invariance in the direction perpendicular to the square objects, so it would come to the same if we studied bars of square cross-section. Under these conditions, the equivalent diffraction patterns for optics are cardinal sines and the contrast between two objects 80% (see Fig. 16.14). Here we consider only square objects with one face in contact with the air (case 2 of Sect. 16.4). There are two types of parameter that can influence resolution: the Young modulus of the object and the matrix, and the size of the object.

Figure 16.15 represents the influence of the Young moduli of the objects on the value of the resolution, for different values of the Young modulus of the matrix. We observe that, when the Young modulus of the objects is small compared with that of the matrix, the resolution is perfect. This is easy to understand. It seems that we image correctly only if the objects are 'soft'. Indeed, they will deform mainly on the side where they are free to move, viz., the air side. So even if the two squares touch one another, the expansion will always allow them to be separated.

Regarding the opposite situation, when the Young modulus of the objects is much bigger than that of the matrix, the resolution reaches a constant value. This can be understood as follows. When the two squares come closer together, the matter between them will be compressed and will thus tend to move up into the air. This in turn will tend to reduce the contrast between the objects. Note that the resolution curves have a similar shape if we normalise with respect to the Young modulus of the matrix. Under these conditions, the following simple analytic form is found for the resolution:

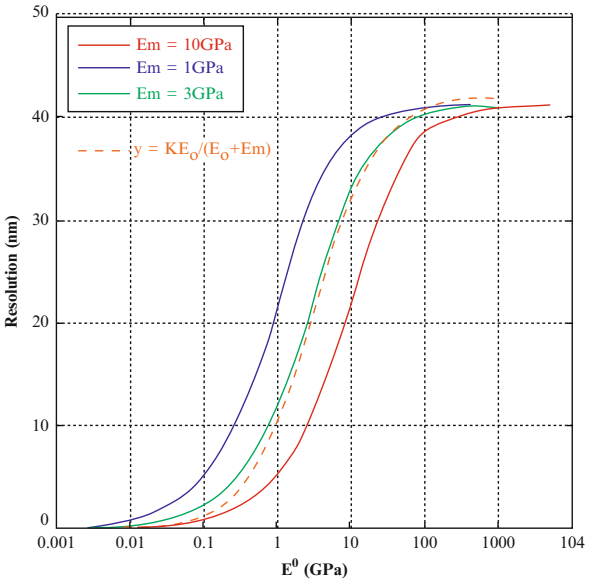

Fig. 16.15 Dependence of the resolution of two buried square objects on their Young modulus for three different values of the Young modulus of the matrix: 1 GPa (*blue*), 3 GPa (*green*), 10 GPa (*red*). The *dashed orange curve* is the constitutive equation

$$\text{resolution} = K\frac{E_\text{o}}{E_\text{o}+E_\text{m}}\,, \tag{16.41}$$

where E_o and E_m are the Young moduli of the object and matrix, respectively, and K is a constant. Note that this fits well with the calculated curve (see Fig. 16.15).

Figure 16.16 shows the effect of the size of the squares on the resolution when $E_\text{o} \gg E_\text{m}$. Note that the dependence is linear with slope 0.4. In this way, the constant K in (16.41) can be replaced to obtain the final expression

$$\text{resolution} = 0.4s\frac{E_\text{o}}{E_\text{o}+E_\text{m}}\,, \tag{16.42}$$

where s is the side of the square.

In most studies, the Young moduli of the objects and the matrix are rather close in value. The resolution will thus be equal to the size of the object divided by 5 (when the Young modulus of the matrix is equal to that of the object). Once again, it should be noted that the resolution does not depend on the wavelength of the illumination and that, for buried objects of side 1 μm, the resolution will be 200 nm. Clearly this resolution is likely to diminish if the objects are buried more deeply.

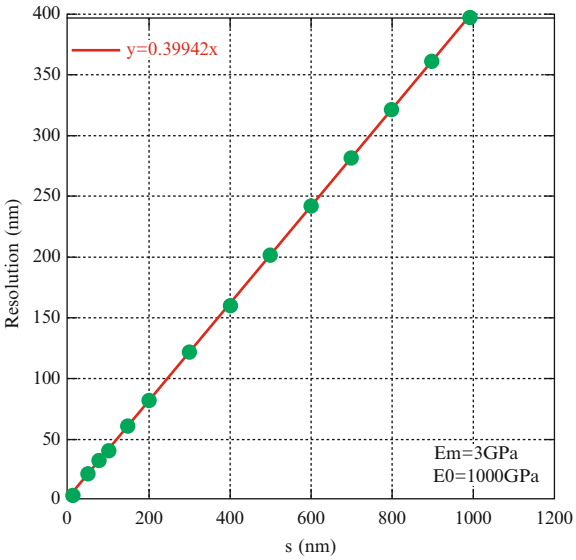

Fig. 16.16 Dependence of the resolution of two square objects on their size *s* (*green points*). Representation of the corresponding linear regression (*red line*)

16.9 Experimental Illustration

The PTIR technique has been used in our research center (Laboratoire de Chimie Physique, Orsay, France) for three years now [16]. This experiment was set up and run at the Centre Laser Infrarouge d'Orsay (CLIO) and now constitutes a line of research in its own right. CLIO is run in a rather unusual way, which allows us to offer our systems to outside users, from other research groups. It is in this context that we have been able to collaborate on several projects in different areas, particularly in biology [17–19], but also in nanophotonics [20]. To illustrate the main features of the PTIR technique by a range of experimental results, we have chosen to present here only the results obtained on biological samples. All the analyses described below have been carried out in air, which means that samples have to be dried. Although one might not think it, biological entities like bacteria, cells, or yeasts, once dried, are not soft and can easily be imaged in contact mode AFM.

16.9.1 Candida Albicans

This work was carried out in collaboration with CHU Reims. *Candida albicans* is a fungus (see Fig. 16.17) and parasite of humans, which may cause mycoses. It is generally found in mucus. Most of the time it is not fatal, and is easily eliminated by the organism. However, in certain kinds of pathology (immunodeficiency), this

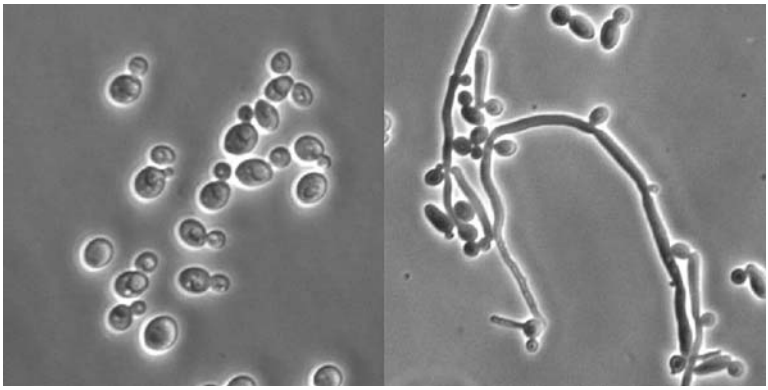

Fig. 16.17 Optical images of *Candida albicans* for two morphologies: blastospores (*left*) and hyphae (*right*)

fungus can become fatal for its host, because it is not repressed by the immune system [21, 22]. *Candida albicans* occurs in two forms in nature: a round form called a blastospore and a long form called a hyphae. Some studies have shown that the long form is responsible for the invasive nature of this fungus [23]. It is thus particularly important to study its virulence, in order to optimise antifungal treatments. In this context, we thus investigated a particular infrared absorption band of this fungus, the glycogen band (1 080 cm^{-1}). Glycogen is a glucose polymer, generally used to store energy in living beings. The distribution of this sugar thus informs as to how the fungus evolves and how it uses its energy reserves for growth [24]. All the studies were carried out in air, i.e., with 'dead' samples.

16.9.2 Escherichia Coli and Its Bacteriophage T5

This work was done in collaboration with the Laboratoire de Physique du Solide in Orsay, France. *Escherichia coli* is a bacterium (see Fig. 16.18) which has long lived in symbiosis with our body, since it inhabits the inner lining of the intestine. This bacterium is not usually pathogenic, but when taken out of its normal environment, it can react violently by secretions that cause irritation (urine infections). This bacterium is rather easy to use. It is generally cultivated in biology labs and used as a DNA factory.

We were thus interested in one of its phages (destroyers) called T5, shown in Fig. 16.19, to try to identify the kinetics of the development of this virus inside the bacterium. Infection by T5 occurs in several well defined stages [25]. In the first of these, the T5 virus finds a site at which to anchor itself on the membrane of the bacterium. Once the three molecules (see Fig. 16.19) at the end of its tail have located their receptors, they lock onto the membrane. This action produces a stress that instantaneously drives the tip of the tail of the virus (shown in yellow in Fig. 16.19)

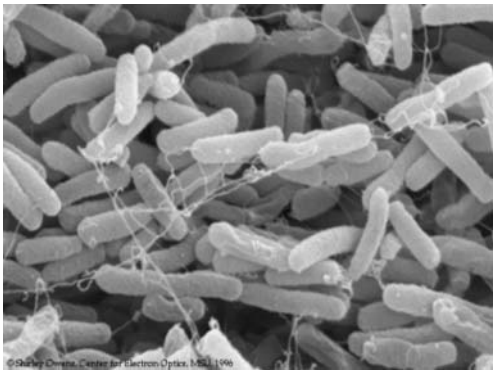

Fig. 16.18 Electron microscope image of *E. coli* bacteria

into the membrane. The end will thus break inside the cytoplasm of the bacterium to release the strand of viral DNA that was stored in the head of the virus. In this process, the capsid (head and tail) act like a nanometric syringe. In the second stage, the DNA machinery of the bacterium is used to duplicate the viral DNA. In this way, viral proteins are generated. They will be used to build new capsids and also to destroy the DNA of the bacterium. Finally, when the viral proteins are assembled into a capsid and the viral DNA strand has been replicated several times, the last stage involves packaging the DNA strands inside the capsids. When all the viruses are filled, the lysis of the bacterium will begin (destruction of the membrane), whereupon hundreds of new viruses will be released into the surrounding medium. Our first studies aimed to investigate *E. coli* alone, then the possibility of detecting the virus in the medium. As for the studies of *Candida albicans*, analyses of *E. coli* and T5 were also carried out in air.

16.9.3 Ultralocalised Infrared Spectroscopy

Experimentally, the absorption is not found by measuring the maximal amplitude of the oscillations (see Sect. 16.7), but rather by a Fourier analysis of the time signal. The Fourier analysis is much more relevant because it exhibits all the contact resonance modes, and it is possible to observe shifts in the modes due to changes in the tip contact conditions (see Sect. 16.6). Rather than measuring $z_{\max}$, it is better to estimate the integral of the mode directly. Before carrying out the spectroscopy, a topographic image of the surface is made. When the biological samples are dried, they are hard enough to be able to image them in contact mode. This image is used to select the study region and position the AFM tip directly in contact with the sample. Then the changes in the integral of the contact mode are recorded as a function of the CLIO laser frequency.

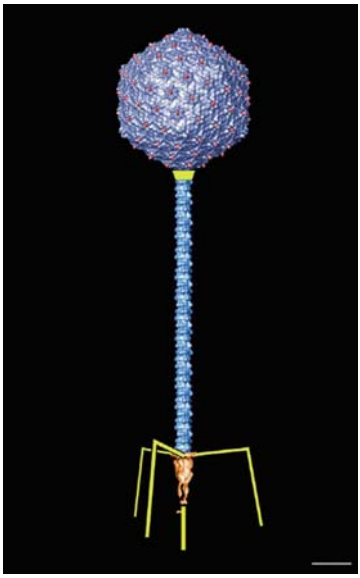

Fig. 16.19 Three-dimensional model of the bacteriophage T5 (IBBMC, Université Paris-Sud, Orsay, France)

Candida Albicans

Samples of *Candida albicans* hyphae were deposited on the surface and several spectra of the glycogen band were obtained at different points of the fungus, i.e., the head, the middle of the tail, and the end of the tail (see Fig. 16.20). The different spectra are found to be rather similar, without any real spectral differences. This absence of any difference suggests that there is no specific development at the end of the hyphae, although one might expect otherwise. It can also be observed that these spectra are all narrower than the spectrum of the glycogen band found by FTIR. However, these results show that the PTIR technique can indeed measure an infrared absorption spectrum with high accuracy, and on a single individual that is smaller than the wavelength.

Escherichia Coli

The infrared spectra obtained on *E. coli* were used as a test to check that our system was indeed well suited to carry out the absorption measurement. Figure 16.21 shows a PTIR spectrum in red, with the FTIR spectrum in green. We observe that the spectra coincide very closely and that the PTIR measurement describes the various bands fairly well: amide I (1550 cm^{-1}), II (1650 cm^{-1}), and the DNA phosphate band (1050 cm^{-1}). The key feature of these results is that the PTIR spectrum was obtained on a single bacterium, while the FTIR spectrum required millions of them.

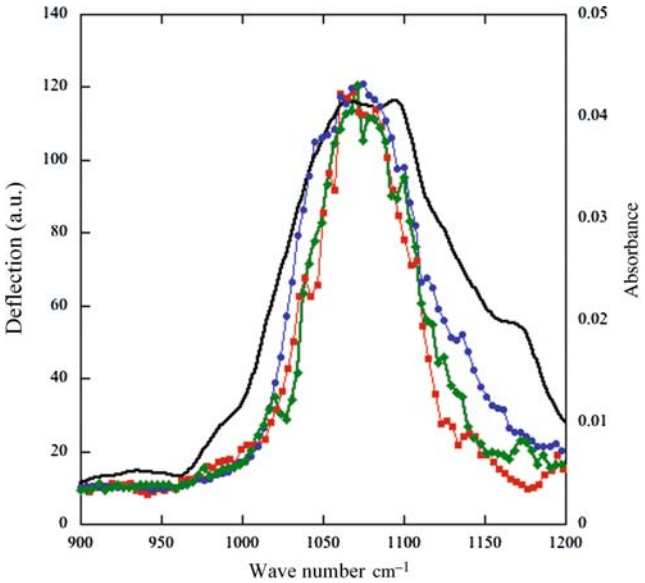

Fig. 16.20 Infrared absorption spectra from different regions of a *Candida albicans*: head (*blue curve*), middle of the tail (*green curve*), and end of the tail (*red curve*), with the FTIR absorption spectrum (*black curve*)

Bacteriophage T5

Once a T5 virus had been located, we were thus able to obtain a spectrum of its (DNA) phosphate band. Figure 16.22 shows a comparison between this phosphate band of the virus and the phosphate band obtained by conventional FTIR spectroscopy. Once again, we observe that the PTIR spectrum reproduces the conventional infrared band very well. This shows that, using our technique, infrared spectroscopy is indeed possible on nanometric objects.

These examples confirm that the contact mode expansion measurements and the infrared absorption by the sample are indeed proportional, as predicted by the simple model in Sect. 16.7. The great advantage here is thus that the measurement is direct rather than differential, i.e., if the the object does not absorb, the PTIR signal is zero, whereas with an optical method, one always requires a variation, often slight, in a nonzero signal.

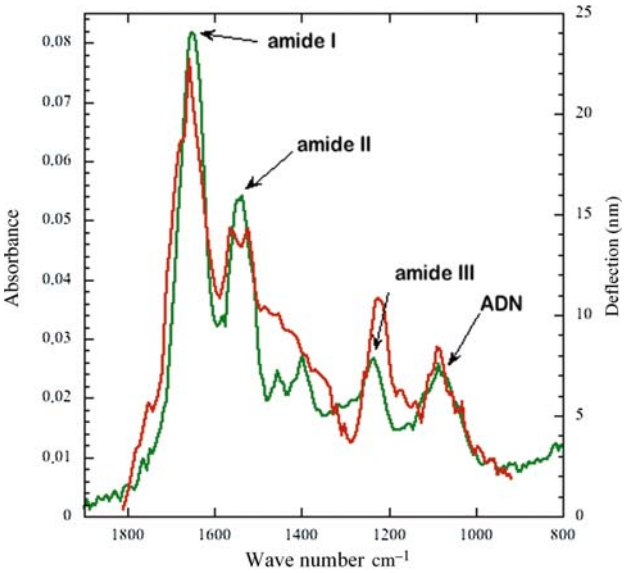

Fig. 16.21 Absorption spectra of a bacterium obtained by PTIR (*red curve*) and FTIR (*green curve*)

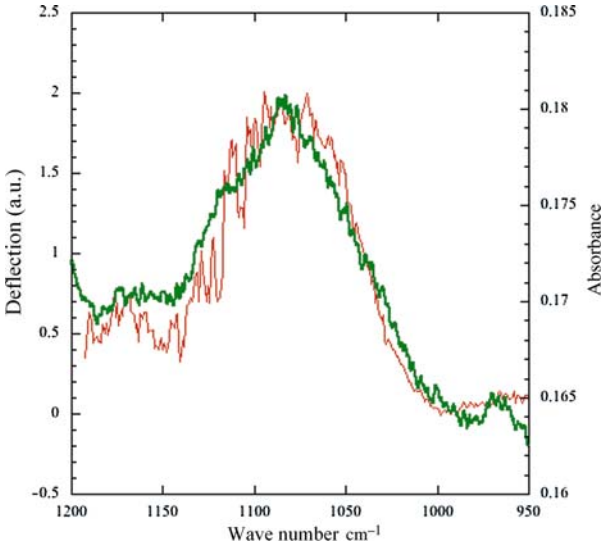

Fig. 16.22 PTIR absorption spectrum of the phosphate band of a T5 virus (*red curve*), compared with the corresponding FTIR spectrum (*green curve*)

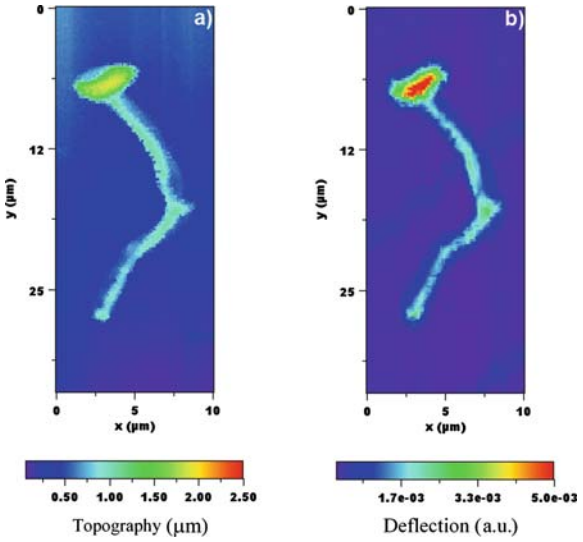

Fig. 16.23 Topographic image of a *Candida albicans* hypha (**a**) and chemical map of the glycogen band of a *Candida albicans* (**b**)

16.9.4 Chemical Mapping at the Nanoscale

Candida Albicans

A map of the glycogen absorption distribution was made by fixing the CLIO laser frequency at 1 080 cm^{-1} and scanning the sample surface (see Fig. 16.23). Note that the chemical map of the glycogen reproduces the height variations of the *Candida*. This proportionality shows that the glycogen is indeed distributed uniformly throughout the sample, and there is no specific storage area. This does not mean that the hypothesis of growth occurring specifically at the end of the tail is necessarily invalid. In biology one has to deal with samples that are extremely reactive to their environment, so it is advisable to repeat the setup many times to guarantee reproducibility of the phenomena. It is therefore our intention to repeat these experiments several times in order to validate our conclusions.

The wave number 1 080 cm^{-1} corresponds to a wavelength of about 10 µm. If we had wanted to study the fungus using confocal microscopy, then given the wavelength, the hole in the diaphragm would have had to have been at best 10 µm. The image of the *Candida* would then have been the size of 1 pixel. In our experiment, the chemical image of the *Candida* is represented by a window of 100×100 pixels, smaller than 10 µm, illustrating once again the sub-wavelength resolution.

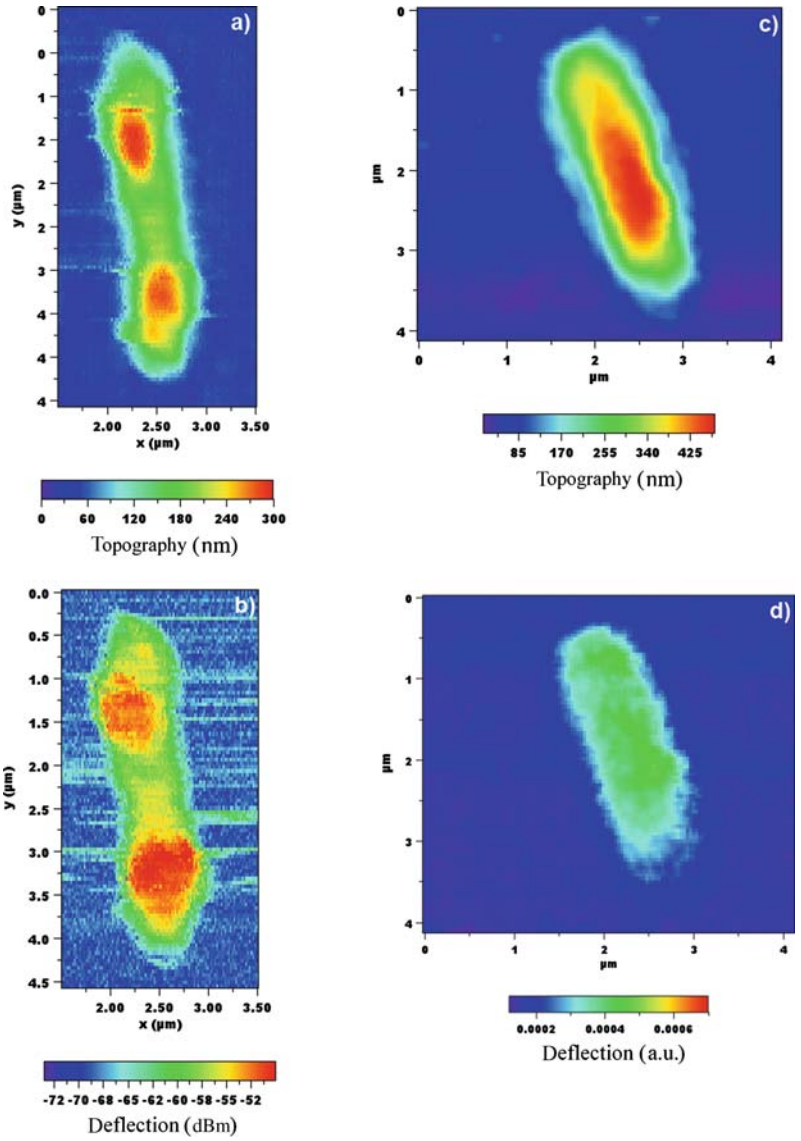

Fig. 16.24 (**a**) and (**c**) Topographic images of an *E. coli* bacterium. (**b**) Chemical map of amide I corresponding to topographic image (**a**). (**d**) Chemical map of the phosphate band corresponding to topographic image (**c**)

Escherichia Coli

Our investigations of the bacterium *E. coli* involved several chemical maps for different bands. Figure 16.24 shows the topography and chemical cartography of amide I and DNA phosphate. We observe that the amide I absorption band is homogeneous.

This seems reasonable given that this band occurs in all the proteins of the bacterium and the proteins are distributed uniformly. As for the phosphate band, which is the signature of DNA, the DNA strand of the bacterium is completely spread throughout the whole bacterium, because there is no nucleus (prokaryote). Furthermore, the phosphate bands are also present in phospholipids in the *E. coli* membrane. All these factors cause the absorption signal to be homogeneous throughout the bacterium.

Note also that the absorption signal is perfectly zero outside the bacterium, which suggests that the absorption signal is identical with that of the topography. As a consequence, there is no broadening of the object due to the expansion, as predicted in Sect. 16.8.1. When mapping isolated objects on a surface, the resolution is indeed dominated by the intrinsic resolution of the AFM.

Bacteriophage T5

We first made maps of the amide I band of T5 on a ZnSe surface (see Fig. 16.25). The resolution of these images is not very good, but it is nevertheless possible to make out the absorption signal of the virus. These viruses are about 90 nm in length. It is no surprise then that the expansion signal is rather weak, and hence difficult to distinguish from the background. The apparent size of the virus in topography is bigger (150 nm) than its normal size. This is due to the convolution of the image by the AFM tip, which was 50 nm in this case. Note that this apparent size is the same as the size of the absorbing region. Once again, we see that it is not the heat expansion that limits resolution, but rather the AFM tip.

Figure 16.26 shows a chemical map of the DNA in a T5 virus actually inside a bacterium. Detection of the virus is perfectly clear. The zoomed image shows that the absorption region is spread over 200 nm. In this case, the apparent size is mainly due to the elastic properties of the bacterium. Indeed, since the contact area between the tip and sample is of nanometric order, it is no longer the convolution of the tip that limits resolution.

As can be seen from the model shown in Fig. 16.27, the apparent size of a buried virus is roughly twice its true size, in agreement with experimental results (see

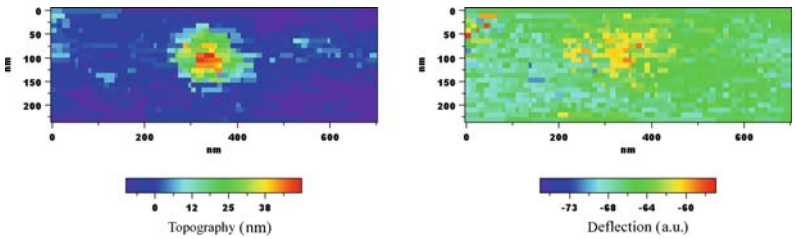

Fig. 16.25 Topographic image of a T5 virus (**a**) and corresponding chemical map of the amide I (**b**)

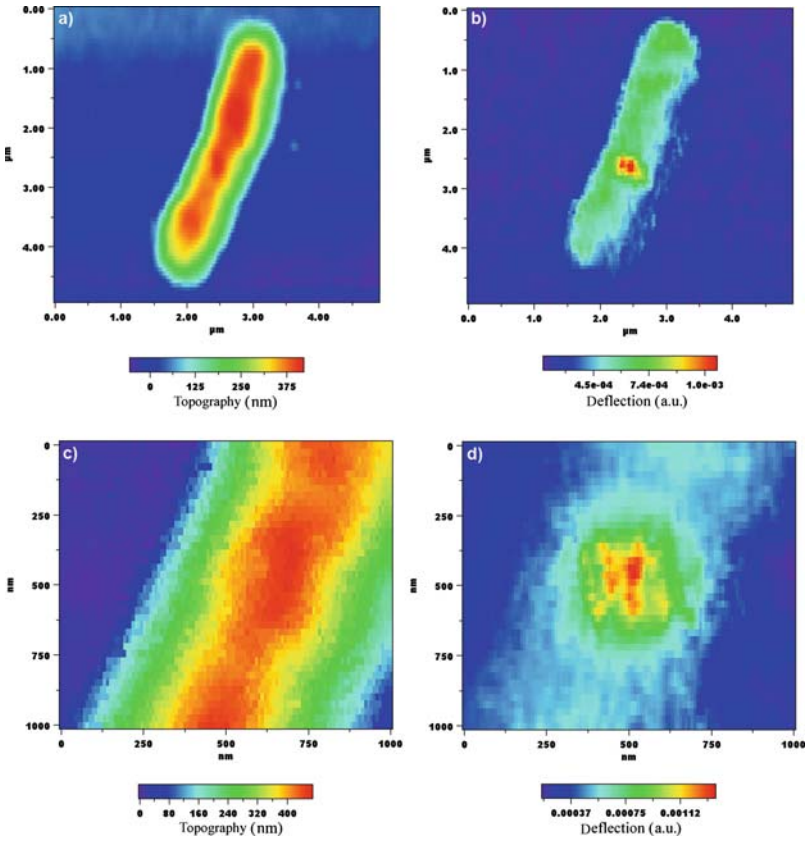

Fig. 16.26 Topographic image of an *E. coli* bacterium (**a**) and corresponding chemical map of the phosphate band (**b**). Magnified topographic image of the region where the virus is buried (**c**) and corresponding chemical map of the phosphate band (**d**)

Fig. 16.26). The possibility of detecting a buried, absorbent nanometric object demonstrates the great sensitivity of this technique, and this will be put to use in our future investigations of the infection of bacteria.

Preliminary calculations have confirmed that, if the object remained close to the surface, the resolution of the chemical imaging would not be degraded. Naturally, when the objects are deeply buried within another object, it will be difficult to detect them because the deformation produced by the photothermal effect at the surface will be diluted. In this sense, PTIR does not escape the similar constraint on near-field techniques, which can only probe regions close to the surface.

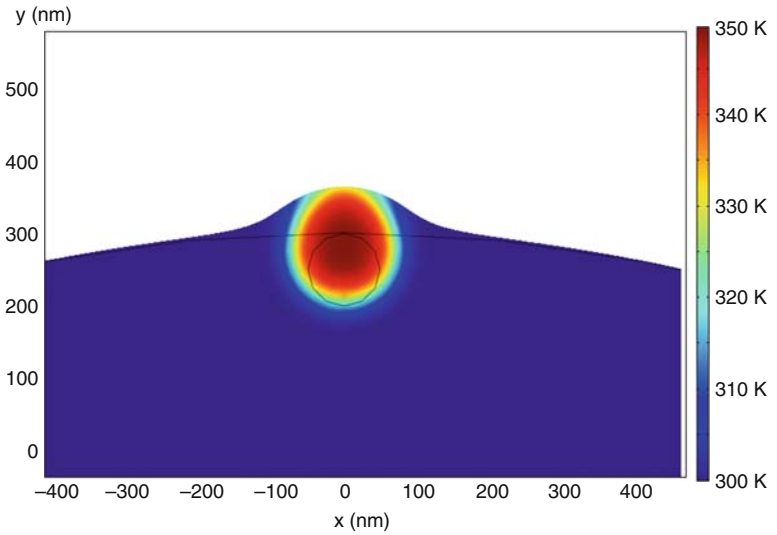

Fig. 16.27 Model of the expansion of the buried T5 virus. The colour code shows the temperature distribution: *red* 350 K, *blue* 300 K. The expanded image is magnified by a factor of two to show the deformations more clearly

16.10 Conclusion

This chapter describes a novel technique for photothermal detection. We have shown that, using an AFM, it is possible to measure the optical properties of materials, and more particularly, to carry out infrared spectromicroscopy. We have demonstrated that the resolution of the PTIR technique is of the same order of magnitude as that of conventional AFM, and this whatever the wavelength used to measure the absorption. The sensitivity of the technique is highly satisfactory since individual objects with sizes around ten nanometers can be identified chemically. The temperature rises required for detection are of the order of a hundred mK, so samples can be irradiated with low laser power. The method using AFM in contact mode and analysing the vibrations of the free modes is not limited solely to excitation by the photothermal effect. PTIR can thus be generalised to any modification of the surface under the tip, provided that it is faster than the normal period of the cantilever.

Regarding the prospects for this type of technique, we have decided to move toward infrared spectromicroscopy on single living cells, either in vitro or in vivo. Indeed, the resolution obtained is sufficient to be able to explore the interior of small cells and bacteria, which is impossible in far-field microscopy. We thus aim to adapt the technique to liquid media, and the first attempts have already given encouraging results [26].

Acknowledgements The following have contributed to this work: R. Prazeres, C. Mayet, F. Glotin, and J.M. Ortega of the Laboratoire de Chimie Physique, Université Paris-Sud, M. De Fru-

tos of the Laboratoire de Physique du Solide, Université Paris-Sud, Orsay, G. Sockalingum of the Unité MéDIAN, UFR Pharmacie, Université de Reims, and D. Toubas of the Laboratoire de Parasitologie-Mycologie, CHU Reims.

References

1. W.D. Perkins: J. Chem. Educ. **63**, A5 (1986)
2. W.D. Perkins: J. Chem. Educ. **64**, A269 (1987)
3. W.D. Perkins: J. Chem. Educ. **64**, A296 (1987)
4. P. Dumas, G.L. Carr, P. Williams: Analysis **1**, 68 (2000)
5. N. Guilhaumou, P. Dumas, G.L. Carr, and G.P. Williams: Applied Spectr. **52** (8), 1029 (1998)
6. N. Jamin, P. Dumas, J. Montcuit, W.H. Fridmann, G.L. Carr, and G.P. Williams: *Accelerator-Based Infrared Sources and Applications*, ed. by G.P. Williams and P. Dumas, SPIE **3153**, 133 (1997)
7. R. Bhargava, I. Levin (Eds.): *Spectrochemical Analysis Using Infrared Multichannel Detectors*, Blackwell Publishing (2005)
8. E. Levenson, P. Lerch, M.C. Martin: Infrared Phys. and Techno. **49**, 45–52 (2006)
9. A. Piednoir, F. Creuzet, C. licoppe, J.M. Ortega: Ultramicroscopy **57**, 282 (1995)
10. D. Palanker, G. Knippels, T.I. Smith, H.A. Schwettman: Opt. Commun. **148**, 215 (1998)
11. A. Cricenti, R. Generosi, P. Perfetti, J.M. Gilligan, H. Tolk, C. Coluzza, G. Margaritondo: Appl. Phys. Lett. **73**, 151 (1998)
12. B. Knoll, F. Keilmann: Nature **399**, 134–137 (1999)
13. N. Gross, A. Dazzi, J.M. Ortega, R. Andouart, R. Prazeres: Eur. Phys. J. Appl. Phys. **16**, 91 (2001)
14. M. Brehm, T. Taubner, R. Hillenbrand, F. Keilmann: Nanoletters **6** (7), 1308–1310 (2006)
15. R. Carminati, J.J. Greffet: J. Opt. Soc. Am. A **12**, 2716–2725 (1995)
16. A. Dazzi, R. Prazeres, F. Glotin, J.M. Ortega: Opt. Lett. **30**, 2388 (2005)
17. A. Dazzi, R. Prazeres, F. Glotin, J.M. Ortega: Infrared Phy. Techno. **49**, 113 (2006)
18. A. Dazzi, R. Prazeres, F. Glotin, J.M. Ortega: Ultramicroscopy **107** (12), 1194–120 (2007)
19. A. Dazzi, R. Prazeres, F. Glotin, J.M. Ortega, M. Alsawaftah, M. De Frutos: Ultramicroscopy **108**, 635–641 (2008)
20. J. Houel, S. Sauvage, P. Boucaud, A. Dazzi, R. Prazeres, F. Glotin, J.M. Ortéga, A. Miard, A. Lemaître: Phys. Rev. Lett. **99**, 217404 (2007)
21. P. Eggimann, J. Garbino, D. Pittet: Lancet. Infect. Dis. **3**, 685–702 (2003)
22. M.A. Pfaller, D.J. Diekema: Clin. Microbiol. Rev. **20**, 133–163 (2007)
23. H. Lui: Int. J. Med. Microbiol. **292**, 299–311 (2002)
24. I. Adt, D. Toubas, J.M. Pinon, M. Manfait, G.D. Sockalingum: Arch. Microbiol. **185**, 277–285 (2006)
25. M. Zweig, H.S. Rosenkranz, C. Morgan: J. Virol. **9**, 526–543 (1972)
26. C. Mayet, A. Dazzi, F. Allot, R. Prazeres, F. Glotin, J.M. Ortega: Opt. Lett. **33**, 1611–1613 (2008)

Chapter 17
Scanning Thermal Microscopy with Fluorescent Nanoprobes

Lionel Aigouy, Benjamin Samson, Elika Saïdi, Peter Löw, Christian Bergaud, Jessica Labéguerie-Egéa, Carine Lasbrugnas, and Michel Mortier

Abstract Luminescence is light emission by materials after absorption of energy. Today, this effect has been made into a very powerful way of characterising materials in a whole range of different fields, across physics, biology, and chemistry. In terms of technological applications, luminescence is also on the point of replacing incandescence for short-range lighting purposes, e.g., pocket lamps, with the advent of white light-emitting diodes. In this chapter, we shall describe a specific application of luminescence to the development of thermal nanosensors. There are four sections. In the first two, we simply describe the luminescence phenomenon, along with several light-emitting materials used for thermometric measurements. In particular, we shall explain how the temperature of a material can be determined from data concerning its luminescence. In Sect. 17.3, we shall discuss the technique of scanning thermal microscopy with fluorescent nanoprobes, together with the experimental setup. In the last section, we shall discuss applications of this technique to image microelectronic devices. The characteristics of the probes and their advantages and disadvantages as compared with other near-field probes will be described in some detail.

17.1 Luminescence

17.1.1 Introduction to Luminescence

Luminescence is an effect that has been known and understood for a long time now. The material, which can be a molecule or a solid, must first be excited by an external source. This excitation depopulates the ground state and fills various excited states, whereupon the return to the low energy state occurs by light emission. There are several ways of exciting materials, such as electroluminescence or photoluminescence. Electroluminescence takes place in electroluminescent diodes and laser

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118 505
DOI 10.1007/978-3-642-04258-4_17, © Springer-Verlag Berlin Heidelberg 2009

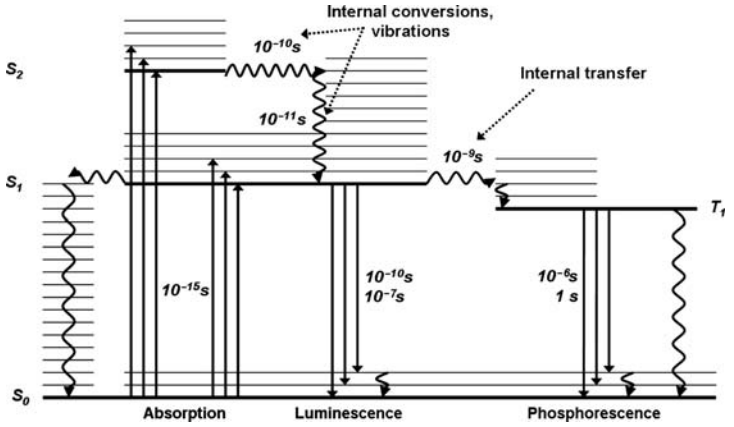

Fig. 17.1 The Jablonski diagram describing the various types of recombination occurring in fluorescent materials

diodes in which the excitation is produced directly by injecting carriers (electrons and holes) into the active part of the material. When this excitation is produced by an electromagnetic wave, which may lie anywhere between the infrared and the ultraviolet, the process is called photoluminescence. The well known Jablonski diagram (see Fig. 17.1) is often used to describe the absorption and relaxation process in a light-emitting system [1].

The Jablonski diagram is self-explanatory. When light has been absorbed, the molecule leaves the ground state S_0 and goes into an excited state S_1, sometimes passing through a higher energy state S_2 on the way. The relaxation from S_2 to S_1, due to vibrations, is very fast, lasting between 10^{-12} s and 10^{-10} s. Once in the state S_1, the molecule can drop back down to its ground state directly by fluorescence, a very fast process lasting between 10^{-7} s and 10^{-9} s. Another possibility is for the molecule to go into another excited state (the triplet state T_1). From this level, direct return to the ground state occurs essentially through non-radiative recombination. It can also occur by light emission if this is allowed by external perturbations such as the crystal field of a matrix, for example. This process, called phosphorescence, is much slower than fluorescence (from 10^{-6} s to a few seconds).

Two quantities play an important role in light-emission experiments: the quantum yield (or quantum efficiency) and the absorption cross-section. The fluorescence quantum yield is the ratio of the number of photons emitted to the number of photons absorbed. For organic molecules, it can reach values of 0.6. The absorption cross-section characterises the absorption efficiency of the molecule. It depends on the excitation wavelength and the type of fluorescing object. It can be considered as the effective cross-sectional area of the molecule which absorbs one photon, and varies between 10^{-15} and 10^{-20} cm^2 for most materials. To calculate the absorption rate of a molecule, one therefore multiplies the absorption cross-section by the incident photon density given in W cm^{-2} or photon s^{-1} cm^{-2}.

The Jablonski diagram in Fig. 17.1 describes the light-emission mechanisms after absorption of photons of higher energy. Materials can also be excited by radiation at energies lower than the emission energy, e.g., by infrared photons. This process is called two-photon absorption luminescence. It is a non-linear mechanism, which can be achieved in two different ways. If the material has no intermediate energy level, as happens in organic molecules, the two photons have to be absorbed simultaneously. The two-photon absorption cross-section is very small, typically of the order of 10^{-45}–10^{-50} cm^4 s/photon,[1] and very powerful lasers must be used, e.g., lasers delivering femtosecond pulses. If the material has intermediate levels, as happens with rare earth ions, a first photon is absorbed, followed by another. This process is known as up-conversion. In contrast to the situation where there are virtual levels, the excitation power does not need to be very high, and low-power continuous wave lasers can be used.

Before describing the materials used in thermometry, we shall briefly discuss the way temperature affects emission processes.

17.1.2 Effect of Temperature on Light Emission

Temperature affects absorption and emission of light in several different ways. The parameters varying with the temperature are the absorption cross-section, the rates of non-radiative recombination and transfer between levels, the lifetime of the excited state, and the energy and width of transition peaks.

The first effect of a rise in temperature is to increase the vibrational motion of the atoms. These lattice vibrations, known as phonons, induce variations in the crystal field which broaden the optical transitions. In addition to line-broadening, higher temperature favours non-radiative recombination from excited states and tends to repress radiative transitions. Fluorescence and phosphorescence will both be affected by an increase in temperature. Phosphorescence may even be much more affected because the population of the triplet state depends directly on non-radiative transfer from the singlet state. Another effect of temperature is to shift the peak energy of optical transitions. This phenomenon can be explained by thermal expansion of the lattice and by electron–phonon interactions. This effect is well known in semiconductors for which the bandgap energy is generally reduced when the temperature rises [2, 3]. Finally, for certain materials containing ions, the relative populations of some neighbouring excited levels depend on the temperature through a Boltzmann distribution. In this case, the relative intensity of two neighbouring emission lines depends on the temperature.

The temperature of a material, or the variations in its temperature, can be determined by analysing one of these quantities, such as the overall intensity of an emission line. However, the sensitivity of the different methods depends significantly on

[1] The two-photon absorption cross-section is the product of two one-photon absorption cross-sections (for instance 10^{-16} cm^2 $\times$ 10^{-16} cm^2) by the time needed for the absorption of the two photons (10^{-15} s). This explains why it is expressed in cm^{4}s/photon.

the material used. Historically, the idea of using light-emitting materials to carry out thermometric measurements goes back to the period 1930–60. The compounds originally used were inorganic and contained Eu^{3+} and Cr^{3+} ions [4]. Later, other glasses or crystals doped with rare earth ions like Er^{3+}, Tb^{3+}, and Ce^{3+} were used. More recently, temperature sensors have been made using organic molecules such as rhodamine B or semiconductor nanoparticles such as CdSe/ZnS (quantum dots).

In the next section, we shall describe in a little more detail how the temperature affects the light-emission properties of some specific compounds, and depending on the material used, we shall indicate the most suitable technique for determining a temperature as accurately as possible.

17.2 Luminescent Materials Used in Thermometry

17.2.1 Organic Molecules. Intensity Variations

Organic molecules have often been used in thermometry. Indeed, they have a high absorption cross-section of the order of 10^{-16} cm^2, so that a high illumination power is not required, and a significant luminescence in the visible. Furthermore, this luminescence is very sensitive to the temperature. Figure 17.2 shows the temperature dependence of the luminescence intensity of rhodamine B between 20°C and 80°C [5, 6]. A reduction of almost 80% is observed, indicating a significant sensitivity to temperature change. Several physical mechanisms are responsible for the drop in fluorescence of organic molecules. In the case of rhodamine B, it seems that a temperature increase induces a structural transformation of the molecule toward a form that no longer emits light [7]. Despite photobleaching problems, rhodamine B molecules turn out to be rather robust, and can go through several cycles of

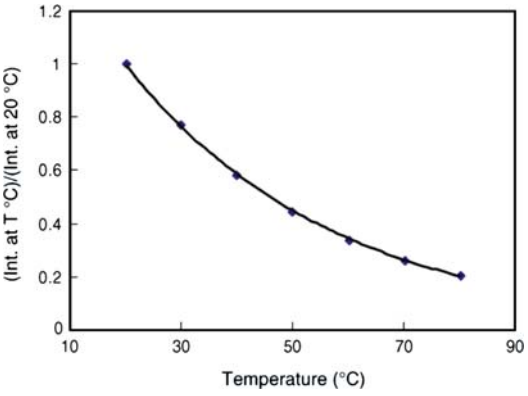

Fig. 17.2 Intensity variations in the fluorescence of rhodamine B at different temperatures. Taken from Arata et al. [6]

temperature increase and decrease before they are fully degraded, with a good level of reproducibility, especially when they are used in a concentrated form. The temperature range nevertheless remains below a hundred degrees or so.

Finally, organic molecules are also good candidates for thermometry because they are relatively easy to handle, in both wet and dry media. They are soluble in certain liquids and can thus be used in microfluidic devices.

17.2.2 Materials Containing Rare Earth Ions

The rare earth atoms occupy the sixth row of the periodic table [8, 9], from cerium to ytterbium, inclusive. Trivalent ions of rare earths are characterised by an unfilled $4f$ shell, protected from external fields by the electrons in the $5s^2$ and $5p^6$ shells. Optical transitions within the $4f$ shell are normally forbidden. However, they may become partially allowed when the ions are located within a matrix, by virtue of the crystal field of the local environment. The fluorescence or phosphorescence lifetimes are relatively long, from a few μs to a few ms, and the emission lines are extremely narrow. Among the rare earth ions, Er^{3+} and Eu^{3+} have often been used in temperature sensors.

Compounds Doped with Er^{3+} and Yb^{3+}. Thermal Equilibrium of Lines

Materials containing Er^{3+} ions have many excited states inducing several light-emitting transitions in the visible (400 nm, 520 nm, 550 nm, 660 nm) and in the near infrared (975 nm, 1 550 nm). Visible transitions can be observed after an excitation in the ultraviolet or the blue (so-called Stokes excitation) or after an excitation in the near infrared (anti-Stokes excitation). The latter, generally close to 980 nm, involves the absorption of two photons [10]. If the material contains only Er^{3+} ions, the photons are either absorbed one after the other by the same ion, or transferred from one Er^{3+} ion to a neighbouring ion. If the material is codoped with Yb^{3+} ions, infrared photons are mainly absorbed by the ytterbium, which has a broad absorption band around 980 nm. The photons are then transferred to the neighbouring Er^{3+} ions in the matrix in an extremely efficient way. The latter excitation process is depicted in Fig. 17.3.

Once excited, the Er^{3+}/Yb^{3+} material has two specific visible transitions (near 520 nm and 550 nm) which are in thermal equilibrium. At a given temperature T, the intensities $I_{520}(T)$ and $I_{550}(T)$ of the two levels are given by

$$I_{520}(T) = p_{520}(T)N_{520}(T)\,, \qquad I_{550}(T) = p_{550}(T)N_{550}(T)\,, \tag{17.1}$$

where the parameters p_i are related to the radiative recombination rates and N_i are the populations of the two levels. If the levels are in thermal equilibrium, then N_{520} and N_{550} are related by

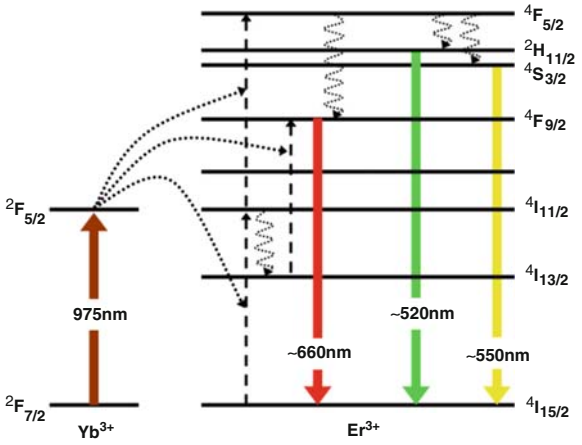

Fig. 17.3 Energy diagram of a material codoped with Er^{3+} and Yb^{3+} ions

$$\frac{N_{520}}{N_{550}} = \exp\left(-\frac{\Delta E}{kT}\right) , \tag{17.2}$$

where ΔE is the energy difference between the two levels, and k is the Boltzmann constant. If we assume that the parameters p_i of the different levels have the same temperature dependence, the ratio of the integrated photoluminescence intensities becomes

$$\frac{I_{520}}{I_{550}} = C\exp\left(-\frac{\Delta E}{kT}\right) , \tag{17.3}$$

where C is a constant depending on the material used. Hence, given the intensity ratio of the two lines, one obtains an absolute measurement of the temperature. This method is based only on the validity of the Boltzmann distribution. Figure 17.4 shows the temperature dependence of the photoluminescence spectra of PbF_2 nanoparticles containing Er^{3+}/Yb^{3+} ions. The two lines, due to transitions between the excited states $^2H_{11/2}$ and $^4S_{3/2}$ and the ground state $^4I_{15/2}$, are in thermal equilibrium. Figure 17.4 also shows the dependence of the quantity $\ln(I_{520}/I_{550})$ on the reciprocal temperature. The curve is linear and does indeed confirm (17.3). Fitting the curve, one obtains a value for ΔE equal to 940 ± 40 cm^{-1}, close to the energy separation between the two levels.

Er^{3+}/Yb^{3+} ions form a very useful combination, which has been intensively exploited in the development of optical thermometers [11–14]. These ions can be incorporated into a wide range of crystalline or amorphous matrices, in the form of powders or nanoparticles with sizes ranging from a few nanometers to several microns. Below, we shall see some applications of these materials in temperature imaging.

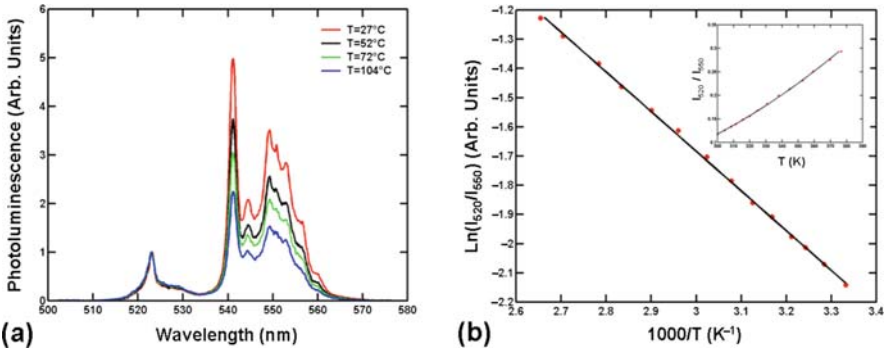

Fig. 17.4 (**a**) Temperature dependence of the photoluminescence spectra of PbF_2 nanoparticles codoped with Er^{3+} and Yb^{3+} ions. The spectra are normalised to their values at 523 nm. (**b**) Logarithmic variation of the ratio I_{520}/I_{550} for the two emission lines as a function of the reciprocal temperature. The curve in the *insert* gives the intensity ratio as a function of T

Compounds Doped with Eu^{3+}. Intensity Variations

Like Er^{3+} ions, Eu^{3+} ions can be incorporated in either a glass or a crystal [15–17], or they can be attached to organic molecules [18, 19]. Under ultraviolet excitation, strong luminescence occurs in the visible, corresponding to transfers between the ^5D_j excited states and the ground sublevels ^7F_i (see Fig. 17.5a) [8]. In many matrices, such as oxides or oxysulfides for instance, emission is influenced by the host levels, or charge transfer states. In general, when the temperature rises, the intensities of many radiative transitions tend to decrease owing to energy transfers toward these levels [15, 20]. However, there are so many possible interactions between the ions and their surroundings that the temperature dependence of the various emission lines is different for each type of host. In some cases, the intensity of certain optical transitions can even increase with temperature [17]. For example, in the case of the compound Y_2O_3:Eu^{3+}, the intensity of the transition $^5D_0 \rightarrow {}^7F_2$ increases or decreases depending on whether the system is excited at 580 nm or 488 nm [17].

When Eu^{3+} ions are associated with organic molecules such as thenoyltrifluoroacetonate (TTA), they can also be very efficiently excited after absorption of ultraviolet radiation by the molecule, followed by transfer to the ions. The reduction in luminescence with temperature is then due to a transfer from the $^5D_0$ level to the triplet state of the TTA molecule [21].

In practice, although the relative populations of several excited levels are governed by Boltzmann statistics, materials doped with Eu^{3+} ions have usually been used as temperature probes by simply analysing the changes in luminescence of isolated transitions. For example, the change in intensity of transitions $^5D_j \rightarrow {}^7F_i$ in the compound La_2O_2S:Eu(0.5%) is shown in Fig. 17.5b [4, 15]. A fast drop is observed for the various lines, indicating a good sensitivity to temperature change. Moreover, each line is reduced in a different temperature range, whence this compound can be used in many applications at both low and high temperatures.

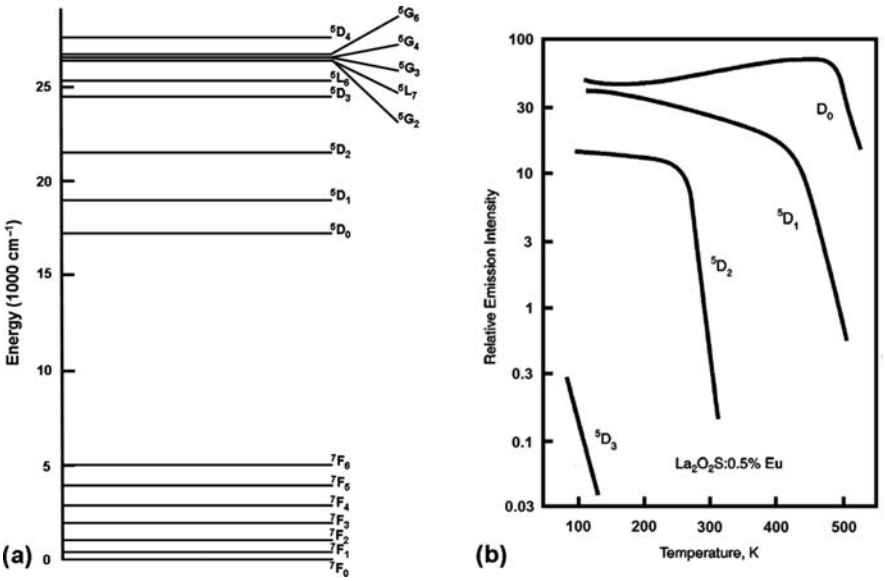

Fig. 17.5 (**a**) Energy levels in Eu^{3+}. Taken from [9]. (**b**) Temperature dependence of ^5D_j transitions in the compound La_2O_2S:Eu(0.5%). Taken from [4] and [15]

17.2.3 Materials Containing Transition Ions. Intensity Variations and Lifetimes

Erbium and europium are not the only ions to have been proposed as temperature sensors. Other rare earths like Ce^{3+} and Nd^{3+} have also been considered for the development of sensitive sensors [4, 22, 23]. Likewise, the optical response of transition ions such as Cr^{3+} in different crystalline matrices (Al_2O_3, LiSAF, YAG) have also been studied in some detail. In the case of Cr^{3+} ions in $LiSrAlF_6$, the drop in luminescence induced by a temperature change can be simply described by considering just two energy levels, as shown in Fig. 17.6a [24]. Starting from the lowest point of the excited state (point I on the $^4T_2$ band), the return to equilibrium at the ground state $^4A_2$ can occur either by light emission, or non-radiatively from the point of intersection between the levels (point Q). The higher the temperature, the more ions will go into the state Q, and the more non-radiative recombinations will occur. According to this simple model, the relaxation rate is given by [24]

$$\frac{1}{\tau} = \frac{1}{\tau_\mathrm{i}} + \frac{1}{\tau_q} \exp\left(-\frac{\Delta E_q}{kT}\right) , \tag{17.4}$$

where τ is the measured fluorescence lifetime, τ_i is the intrinsic radiative lifetime, $(1/\tau_q)\exp(-\Delta E_q/kT)$ is the rate of transfer to point Q, and ΔE_q is the energy separation between points I and Q. From this very simple description, the temperature

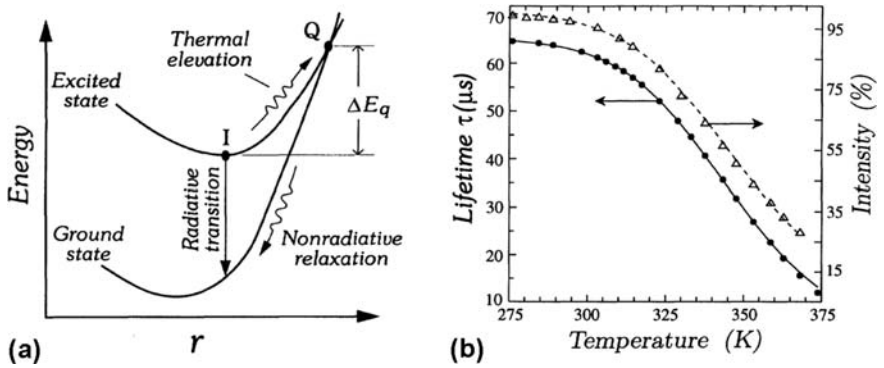

Fig. 17.6 (**a**) Model used to describe the temperature dependence of fluorescence by Cr^{3+} ions in $LiArAlF_6$. (**b**) Fluorescence intensity and lifetime variations measured for an $LiSrAlF_6$ crystal containing Cr^{3+} ions. Taken from [24]

dependence of the luminescence intensity and lifetime variations can be expressed as follows:

$$I = \frac{I_0}{1 + \dfrac{\tau_\mathrm{i}}{\tau_q} \mathrm{e}^{-\Delta E_q / kT}} \tag{17.5}$$

and

$$\tau = \frac{\tau_\mathrm{i}}{1 + \dfrac{\tau_\mathrm{i}}{\tau_q} \mathrm{e}^{-\Delta E_q / kT}} \,. \tag{17.6}$$

The fluorescence intensity and lifetime variations of Cr^{3+} ions in an $LiSrAlF_6$ matrix have been measured experimentally and the results are shown in Fig. 17.6b. The functional dependence of the intensity and lifetime on the temperature is the same, and the fit obtained using (17.5) and (17.6) shows that the simple model above provides a reasonable description of thermal effects in this material.

17.2.4 Semiconducting Quantum Dots. Intensity and Wavelength Variations

Semiconducting quantum dots such as CdSe/ZnS are currently enjoying considerable development in biology [25, 26]. Like organic molecules, they have a good quantum yield (> 0.2) and emit intense luminescence at wavelengths depending on the size of the dot, with emission ranging from blue to red as the size goes from 2 nm to 10 nm. Photoluminescence from this type of nanostructure varies sensitively with the temperature [27–30]. Figure 17.7 shows the temperature dependence of the luminescence spectra of CdSe/ZnS quantum dots over the range 100–315 K [27].

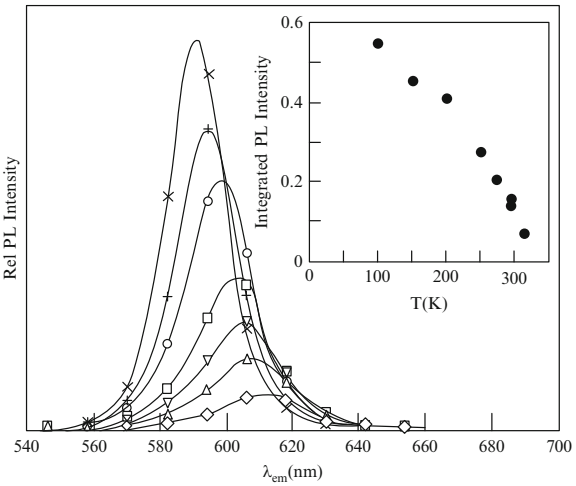

Fig. 17.7 Temperature dependence of the photoluminescence spectra of CdSe/ZnS quantum dots [27]

As the temperature increases, the luminescence intensity decreases, the emission line broadens, and the transition shifts to lower energies. The luminescence intensity varies by about 1%/°C between 250 and 315 K. Around 275–315 K, the drop in intensity is very likely due to the thermal escape of the CdSe carriers to ZnS barriers. The changes in the transition energy obey Varshni's empirical law [2], viz.,

$$E_{\mathrm{g}}(t) = E_{\mathrm{g0}} - \alpha \frac{T^2}{T+\beta}\,, \tag{17.7}$$

where E_{g0} is the bandgap energy at 0 K, α is a constant depending on the material, and β is a parameter close to its Debye temperature. Recent measurements have shown that the drop in the bandgap of quantum dots is close to that of the bulk CdSe material [28]. The variations in the full-width at half-maximum of the luminescence peak can be described by the relation [28]:

$$\Gamma(T) = \Gamma_{\mathrm{inh}} + \sigma T + \Gamma_{\mathrm{LO}} \left(\mathrm{e}^{E_{\mathrm{LO}}/kT} - 1 \right)^{-1}\,, \tag{17.8}$$

where Γ_{inh} is the inhomogeneous broadening due to the change in size of the quantum dots, σ is the acoustic phonon/exciton coupling coefficient, Γ_{LO} is the optical phonon/exciton coupling coefficient, and E_{LO} is the optical phonon energy. Relations (17.7) and (17.8) thus provide a simple description of the effect of temperature on semiconductor photoluminescence spectra.

Regarding the practical uses of this kind of particle, temperature ramping cycles carried out recently [29] have shown a good level of reversibility of these effects, making them good candidates for thermometric measurement, at least for temperatures below 50 or 60°C.

17.3 Development of a Scanning Fluorescent Probe for Temperature Measurements

We set out to develop a scanning probe for local temperature measurements using a single fluorescent particle. The idea is of course to observe the heating of a device with the best lateral resolution possible. The great advantage of a scanning thermal probe is that it is non-invasive. So once it has been characterised, the sample can be reused for the desired applications.

17.3.1 Choice of Material. Reversibility

We chose to use particles containing erbium and ytterbium ions. These particles are either fragments of fluoride glasses, or PbF_2 nanoparticles. Temperature measurements are carried out by comparing the intensities of the two emission lines located around 520 and 550 nm (see Fig. 17.4). There are two reasons for carrying out a measurement using the two luminescence lines. First of all, since the two energy levels are in thermal equilibrium, a comparison between the intensity of the two lines can determine the absolute temperature of the material. Furthermore, a scanning probe device, combined with an optical microscope, is often sensitive to drift. In particular, the fluorescence emission must be detected by a microscope objective placed above the sample. A drift along the z axis of the sample and the sample holder can slightly alter the light gathering conditions and thereby induce artifacts. By measuring the intensity of two lines and comparing their ratio, any drift effect will be cancelled out.

In order to assess the sensitivity of the technique and test reversibility of the measurement, we carried out a far-field measurement of the luminescence by placing a relatively large fluorescent particle (around 2 μm) on a heater consisting of a nickel strip of width 4 μm with an electric current passing through it. Figure 17.8a is a photograph of the strip. The position of the fluorescent particle (a glass fragment codoped with Er^{3+} and Yb^{3+} ions) is indicated by the arrow. The particle and the Ni strip are illuminated by a laser diode at 975 nm. The changes in the fluorescence of the lines at 520 nm and 550 nm were measured simultaneously for different electric currents in the strip. They are shown separately in Figs. 17.8b and c, when the current is ramped up and then down again to 0 mA. While the electric current is being raised (and with it therefore the temperature of the strip, heated by the Joule effect), an increase in the fluorescence of the line at 520 nm and a decrease in the fluorescence of the line at 550 nm are observed. This shows that the two lines are in thermal equilibrium and that the increase in temperature depopulates the $^4S_{3/2}$ level in favour of the $^2H_{11/2}$ level. However, these measurements are sensitive to drift in the device. Indeed, when the current is ramped down, a slight shift is observed in the curves. This shift is not due to degradation of the material, but rather to drift, probably induced by the temperature rise in the strip. The curve in Fig. 17.8d represents

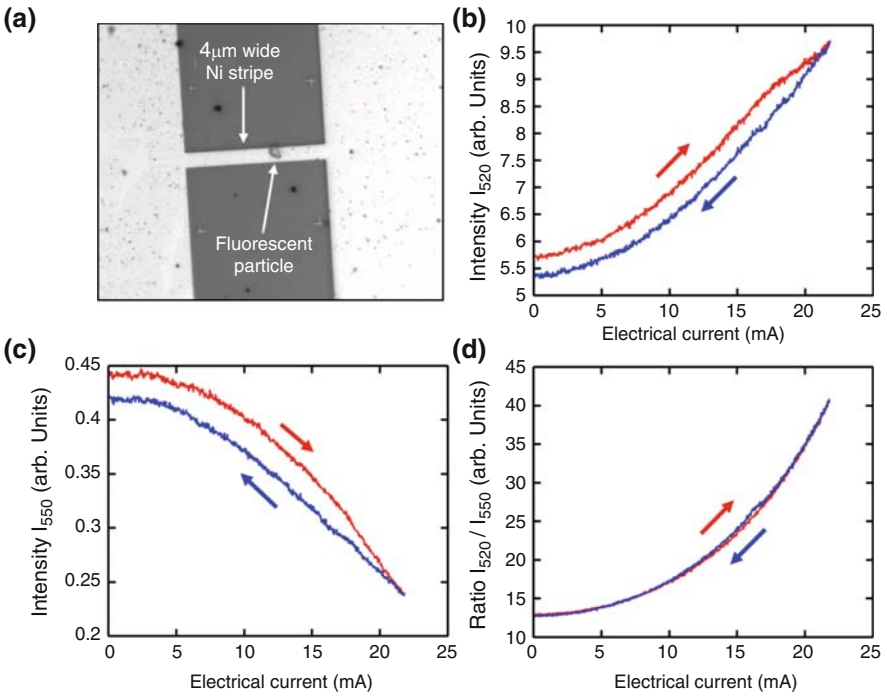

Fig. 17.8 (**a**) Microphotograph (97μm × 77μm) of an Ni strip of width 4 μm. (**b**) Intensity of light emitted at 520 nm. (**c**) Intensity of light emitted at 550 nm. (**d**) Ratio of the intensities of the two luminescence lines. *Red* and *blue curves* are obtained respectively when increasing and decreasing the current in the strip

the intensity ratio of the two lines. Perfect coincidence is observed between the rise and fall of the temperature, showing that the drift effects cancel.

This therefore provides a material and a robust temperature measurement technique, able to operate well above a hundred degrees without risk of photodegradation. These fluorinated materials are not water-soluble and are extremely resistant to acids (very likely down to a pH of unity). The resistance to an alkaline medium is doubtless less good, but not many studies have yet been carried out regarding this point. In the next section, we shall describe how scanning local probes can be made from small fluorescent objects.

17.3.2 Making the Probes

It is always difficult to make a local thermal probe because one requires a thermometric sensor, e.g., a thermocouple, thermoresistive junction, or fluorescent object, at the end of the tip of an atomic force microscope. In our case, the tip carrying

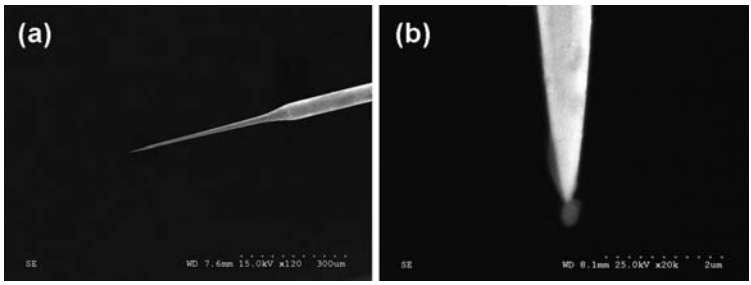

Fig. 17.9 Scanning electron microscope images of a tungsten tip after electrochemical etching (**a**), and after gluing a fluorescent particle on the end (**b**)

the fluorescent object must be highly tapered, because it must not get in the way of the collection of luminescence by the microscope objective placed just above it. It is thus almost impossible to use the usual silicon or silicon nitride cantilevers preferred for atomic force microscopy, simply because their size would block the efficient gathering of emitted light. We thus opted for tungsten tips. The electrochemical fabrication process used for these tips is well known and well understood [31], as a result of research carried out in scanning tunnelling microscopy and near-field optical microscopy.

The highly tapered tungsten tip is obtained after electrochemical etching in a KOH solution (see the scanning electron microscope image in Fig. 17.9a). It is then coated with a polymer that can be crosslinked by exposure to ultraviolet radiation. The deposit is achieved by immersing the end of the tip (to a depth of 1–2 μm) in the polymer using a micro- or nanomanipulation system. Immersion of the tip in the polymer is controlled by a standard optical microscope using an objective with high numerical aperture. The polymer remains viscous until photo-crosslinked. This allows time to stick on the fluorescent object. The gluing stage is carried out by bringing the tip toward a silicon surface on which the nanoparticles have been deposited, using a piezoelectric system for the approach [32]. Figure 17.9b is an image, taken using a scanning electron microscope, of a tip at the end of which a particle of diameter around 200–300 nm has been glued.

The position of the fluorescent object is very important because it must be located very precisely at the end of the tip. The viscosity of the polymer is a critical parameter. If the polymer is too fluid, it sometimes happens that the particle does not stick at the very end, but moves up the tip. It may then be located several tens or even hundreds of nanometers from the end of the tip, which is a significant handicap when making thermal measurements because, in order to obtain good lateral resolution, heat transfer from the surface to the tip must occur by direct contact and not by conduction through the air, or indirectly through the tungsten.

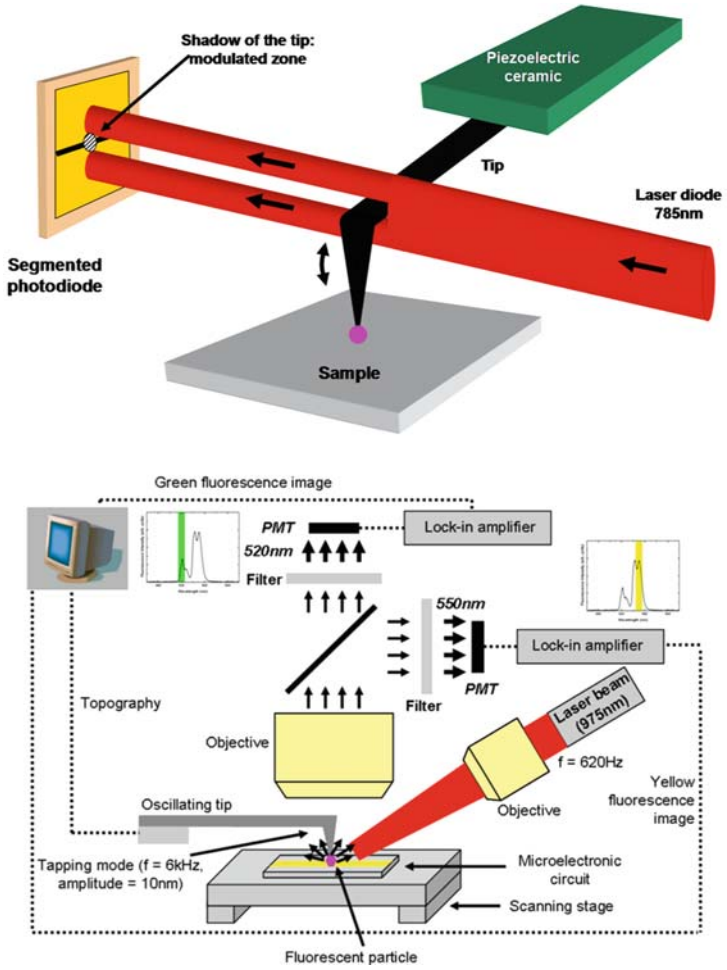

Fig. 17.10 Experimental setup for fluorescent SThM

17.3.3 Experimental Setup for Fluorescent SThM

The experimental setup for fluorescent SThM is rather sophisticated. In addition to the equipment required for other types of scanning probe microscopy, there must be a device to excite and collect fluorescence. Figure 17.10 shows the general experimental setup. The upper part shows the mechanical features of the microscope (atomic force microscope), while the lower part illustrates the optical and thermal features.

In a scanning probe microscope, either the tip–sample distance is held constant, or the tip–sample force is held constant in contact mode. Among the various operating modes available, we use the tapping mode [33]. This is far from being the best

suited to thermal imaging. However, it turns out to be quite convenient in our case, due to the rather special arrangement required for illumination and collection.

The tip is first bent at an angle of 90° at a distance of about 1 mm from the end, as shown in Fig. 17.10. The horizontal part is then placed on a piezoelectric slab and forms the vibrating cantilever. In tapping mode, the cantilever is caused to oscillate at its resonant frequency, between 3 and 10 kHz, and its amplitude of vibration is held constant during image acquisition. To measure this amplitude, usually between 5 and 20 nm, we use a 785 nm laser diode whose beam crosses the horizontal part of the cantilever on its way to a two-segment photodiode. As it oscillates, the tip modulates the laser beam arriving at the two photodiodes, and this produces an alternating signal proportional to the oscillation amplitude of the tip. When contact is made with the surface and the sample surface scanned, the vibration amplitude is held constant using a servosystem (Stanford SRS SIM960).

To excite the fluorescent object placed at the end of the tip, the system is illuminated laterally, focusing the exciting laser diode ($\lambda = 975$ nm) by means of an objective with a long working distance (Nikon SLWD LPlan, $\times 20$, N.A.=0.35). The angle of incidence on the sample surface lies between 10 and 30° from the horizontal, producing a relatively large focusing spot, about 100 μm in diameter. During illumination, the excitation power is kept as low as possible (< 100 mW), so that the incident radiation does not heat the tip or the sample surface. The intensity of the illumination is modulated at low frequency (< 600 Hz).

The light-gathering objective (Olympus LMPlan FL) has numerical aperture 0.8 and a high magnification ($\times 100$). This large numerical aperture allows fluorescence to be collected over a total angle of around 108°. The working distance of the objective is greater than 3 mm, which is very high for this magnification and numerical aperture. This gap allows the tungsten tip to be placed between the surface and the output lens of the objective.

After collection, the fluorescence is split into two arms, the wavelength is selected using two interference filters of passband 10 nm, whereupon it is subsequently guided into two photomultipliers (Hamamatsu). The electrical signals are then sent to two lock-in amplifiers (EGG DSP7260), synchronised on the modulation frequency of the laser at 975 nm. We thus measure the average value of the fluorescence signal while the tip oscillates on the sample surface. A very sensitive multichannel spectrometer has also been connected to the detection device in order to acquire a luminescence spectrum when the tip is located at a precise point on the surface. The temperature of the fluorescent object can then be determined directly. We shall now describe some applications of this technique.

17.4 Applications

Operating microelectronic devices have already been imaged many times using fluorescent objects deposited on the device surface [6,21,34–36]. Very good temperature resolutions have been obtained. Kolodner and Tyson [21,34] were able to measure a

temperature to an accuracy of 0.01°C. Since far-field techniques cannot go beyond the resolution limit of classical optics (typically $\lambda/2$), our aim is to observe hot spots with a lateral resolution well below this limit. The problem is that, in a high-temperature device, heat tends to spread, and it is thus difficult to obtain a sample in which the lateral resolution can really be evaluated. In the following sections, we shall discuss a few examples of temperature visualisation using the fluorescent SThM technique. We first describe d.c. measurements carried out on several types of structure. Then we shall show that the technique can in fact operate when the device is running on alternating current. Finally, we shall show that it is also possible to study heat transfer between tip and sample.

17.4.1 Direct Current Measurements

The luminescence measurements we carry out are above all optical measurements. It is from variations in the optical signal that we determine a temperature. Consequently, since the particle will come into contact with the surfaces of several materials with different dielectric properties, e.g., a metal or an insulator, precautions must be taken during image analysis, because its fluorescence will depend on where it is located. Figure 17.11 shows what causes variations in the light emitted by the particle.

To begin with, if the surface is not uniform, the incident field will not form a homogeneous distribution over the sample. Indeed, the field at the surface results from interference between the incident light and light reflected or scattered by structures at the surface. This leads to a highly inhomogeneous distribution, which will then be 'read' by the particle as it scans. In addition, the fluorescence itself will be influenced by the micro- or nanopatterned surface. So the particle will not emit the same amount of light when located on a metal as it will when located on a glass. On a metal, several things may happen. Generally, the metal will more efficiently reflect or scatter the fluorescence into the microscope objective. However, for very

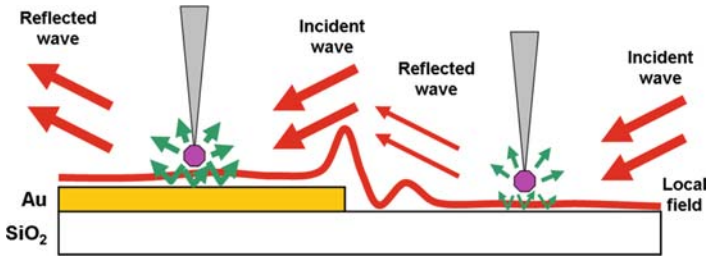

Fig. 17.11 Causes of fluorescence contrast between different materials. In this simple example, there will be more fluorescence on the metal than on the glass, because for one thing the total field at the surface can be greater, and for another the metal reflects more fluorescence into the detector

small particles, or for ions located very close to the surface, their luminescence may actually be reduced by non-radiative transfer to the metal.

During scanning, we measure the intensities of two luminescence lines and take the ratio, thereby cancelling unwanted effects. However, if the materials making up the device have different reflection coefficients at 520 nm and 550 nm, a contrast can be observed between the different zones, even after taking the intensity ratio of the two luminescence lines. We shall now give some examples in which different situations are encountered. In order to observe the thermal effects and eliminate artifacts due to optical contrast, we shall always compare the images obtained with and without current going through the device.

Temperature Rise in a Polysilicon Circuit

The first example concerns a resistive polysilicon circuit of width around 20 μm [37]. The polysilicon is deposited on a silicon oxide substrate and this is then coated with insulating layers of oxide and silicon nitride. Figures 17.12a and b show an optical microphotograph of the whole circuit, together with the topography of the analysed region. Figures 17.12c and d show the ratio I_{520}/I_{550} of the luminescence peaks for two values of the electrical current in the strip ($i = 0$ mA and $i = 50$ mA). It is observed that, for the image obtained without current, there is a contrast between the polysilicon and the silicon. This contrast is not very pronounced (going from 13 to 15), but it shows that the ratio of the reflection coefficients for fluorescence at 520 nm and 550 nm (and the coefficients themselves) is different in the two zones. The presence of several passivation layers at the surface and an insulating layer between the strip and the substrate can also accentuate this contrast through interference phenomena.

The image of the intensity ratio I_{520}/I_{550} is completely different when a current goes through the circuit. To begin with, the scale has changed and the increase indicates a rise in the overall temperature of the structure. On the polysilicon strip, the bends are the hottest zones, because the current density, and hence the Joule effect, is highest there. Note also that the temperature in the upper part of the image, corresponding to the central region of the circuit, is higher than outside the circuit.

However, this image does not precisely represent the temperature, because the contrast depends on the different regions of the sample. To convert to a temperature image, the image obtained must be normalised by a coefficient that depends on each region. To do this, consider the image obtained for zero current. The temperature of the structure is around 30°C. For this temperature, each region is characterised by a value of the ratio I_{520}/I_{550}: 13 for the polysilicon strip and 15 for the silicon substrate. To obtain the temperature of the structure when the current is not zero, we must compare the way the contrast changes in each region. For example, in the arm of the part of the substrate trapped in the polysilicon strip, the contrast is about 28. There is thus a change from 15 to 28, which corresponds, after comparing with the calibration curve (see Fig. 17.4), to a temperature of around 100°C.

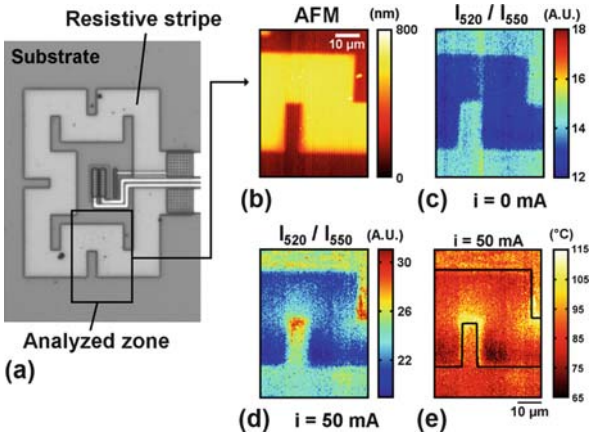

Fig. 17.12 Temperature rise in a polysilicon strip. (**a**) Microphotograph showing the overall structure. (**b**) Topography of the analysed region. (**c**) Image of the intensity ratio I_{520}/I_{550} of the fluorescence lines for $i = 0$ mA. (**d**) As in (**c**) but for $i = 50$ mA. (**e**) Conversion of the optical image in (**d**) to a temperature image. *Black lines* indicate the boundary between the substrate and the polysilicon strip [37]

Figure 17.12e shows the temperature image reconstructed for the whole structure. It has been obtained by defining regions in which the contrasts have been compared and converted into temperature values. The first thing to note is that the whole structure has reached a high temperature (at least in the range 65–75°C). This is due to the large size of the polysilicon strips (width 20 µm) and the presence of several insulating layers at the surface and at the interfaces, which favour diffusion of heat to distances of several tens of microns from the strip. We also note that the reconstruction is not perfect, particularly at certain points on the boundary between regions. Although the temperature agreement is good at the hot spots, it is less good in the lower part of the strip, which seems to be at a slightly lower temperature than the neighbouring substrate, something that is hard to countenance. This first experiment nevertheless provides a good illustration of the different stages required to obtain the temperature image using luminescence lines in thermal equilibrium. We shall now describe a much simpler example, for which there is no contrast between the different materials of the structure when the ratio of the fluorescence images at 520 nm and 550 nm is taken.

Temperature Rise in an Aluminium Oxide Strip

The example to be described here is rather special in that the intensity ratio I_{520}/I_{550} is constant over the whole surface of the device under investigation. It consists of an aluminium strip of width, length, and thickness equal to 1.25 µm, 10 µm, and 50 nm, respectively. The strip was vapour deposited onto an oxidized silicon substrate (oxide thickness $\sim$ 50 nm). During deposition, the aluminium was also

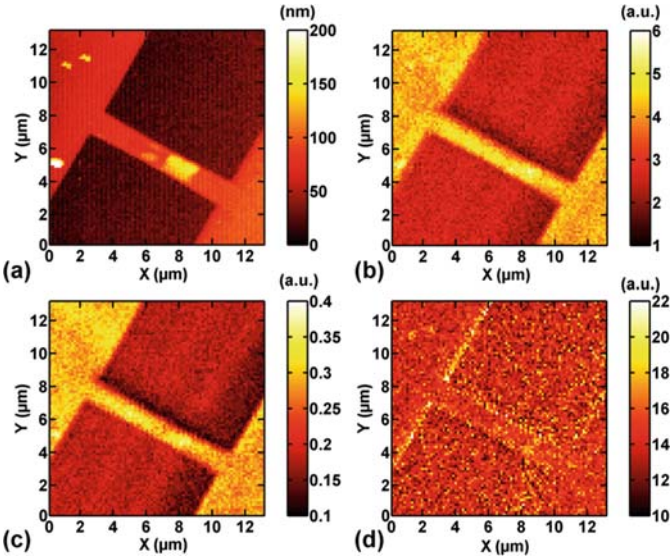

Fig. 17.13 Aluminium strip on an oxidized silicon substrate. (**a**) Topography. (**b**) Fluorescence image at 520 nm. (**c**) Fluorescence image at 550 nm. (**d**) Fluorescence intensity ratio I_{520}/I_{550}. There was no current in the structure [38]

oxidised in order to increase its resistivity and hence cause it to heat up more easily when an electric current passes through it. As for the last example, we began by characterising the structure in zero electric current conditions. The topography, fluorescence images at 520 nm and 550 nm, and the ratio of these two images are shown in Fig. 17.13 [38]. An increase in the intensity of the two fluorescence lines is observed on the aluminium strip. The ratio of these two images (see Fig. 17.13d) only gives a very slight contrast, indicating that the difference in reflection coefficient between the aluminium and the silicon has cancelled out.

An increase is nevertheless visible along certain edges of the structure and in particular on the left-hand side. It may be that this contrast is a shading effect due to the lateral illumination by the laser diode. It may also be an anisotropy induced by the probe, by the more or less conchoidal shape of the fluorescent object, or a shift in its position to one side of the tip. This is therefore an artifact, and poses no threat once identified. Artifacts are commonplace in scanning probe microscopy, and it is a further reminder that a zero-current reference image must always be taken in order to identify genuine thermal effects.

Figure 17.14 shows the current dependence of the contrast I_{520}/I_{550} when a current is passed through the strip. Images are plotted with the same colour scale. An increased contrast is clearly visible on the strip, which shows that, in this region, the intensity of the peak at 520 nm has increased as compared with the peak at 550 nm. This increase reveals a local temperature rise. Unlike the case discussed in the last section, the contrast in this image is directly proportional to the temperature in all regions of the structure.

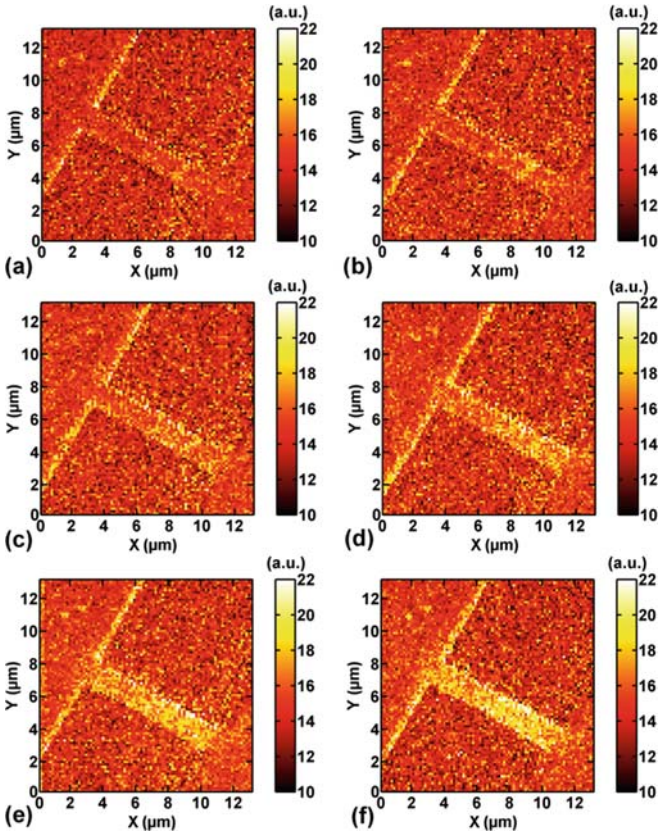

Fig. 17.14 Intensity ratio I_{520}/I_{550} for different values of the electric current through the structure: (**a**) 0, (**b**) 3, (**c**) 5, (**d**) 7, (**e**) 8, (**f**) 9 mA. These values correspond to current densities in the strip in the range 0–$1.44 \times 10^{+11}$ A m^{-2} [38]

The strip temperature reached about 45°C for a current of 9 mA. Beyond 9 mA, the strip is damaged and eventually ruptures. However, a numerical simulation of the temperature acquired by the strip nevertheless showed that the temperature must have been higher [38], implying that the surface–particle heat transfer is not optimal. This effect will be discussed further in Sect. 17.4.3. The lateral thermal resolution of these images (determined from a section taken at right angles to the strip) is around 250 nm, which is slightly less than the size of the fluorescent object used. This resolution is rather good, given the width of the aluminium strip. It can also be explained by the thinness of the silicon oxide layer (50 nm). The silicon substrate absorbs heat efficiently and prevents lateral diffusion into the insulating layer.

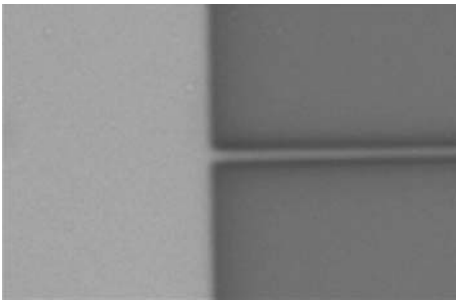

Fig. 17.15 Microphotograph of a nickel strip of width 500 nm, deposited on a silicon oxide substrate. Image size 21 μm × 14 μm

Strips of Submicron Width

We have also characterised metal strips with widths less than the micron, viz., 500 nm and 200 nm. In this case, we are at or below the lateral resolution limit of standard optical microscopy. The structures consist of nickel strips deposited on an oxidised silicon surface (see the microphotograph shown in Fig. 17.15). The thicknesses of the nickel and the oxide layer are 40 nm and 500 nm, respectively.

Figure 17.16 shows fluorescence images at 520 nm and 550 nm, together with their ratio, obtained for current values 0 and 3 mA [39]. The strip investigated here has a width of 500 nm (see Fig. 17.15). Figures 17.16a and b show the fluorescence of the particle when it scans the surface for zero current. Concerning the two lines, we clearly observe greater luminescence on the metal than on the silicon substrate. The ratio of the two images shows a low contrast between the metal and substrate (see Fig. 17.16c). When a current of 3 mA is passed through the structure, the intensity of the line at 520 nm (see Fig. 17.16d) is similar to that obtained at 0 mA. However, a slight increase in the luminescence is observed on the strip. On the other hand, the luminescence at 550 nm (see Fig. 17.16e) drops off significantly on the wire, whereas it remains constant on the contact. Such behaviour is typical for the particles doped with rare earths that we use. In agreement with what was shown in Figs. 17.4 and 17.8, a temperature rise causes a transfer of electrons from the $^4S_{3/2}$ level to the $^2H_{11/2}$ level, and this explains the large drop in intensity of the line at 550 nm and the slight increase in intensity of the line at 520 nm in places where the temperature increases. The ratio of the two lines (see Fig. 17.16f) clearly illustrates the heating of the strip as compared with regions that remain cold.

It is difficult to assess the lateral thermal resolution for observation of this structure. The temperature profile shown in Fig. 17.16f displays a rather slow decrease in the thermal signal. This is a different behaviour to the case discussed in the last section, for which the temperature drop was very sudden in passing from the aluminium strip to the substrate. The main difference here seems to come from the thickness of oxide between the silicon and the metal. The nickel strip lies on a layer of SiO_2 of thickness 500 nm, whereas the aluminium strip is separated from the substrate by

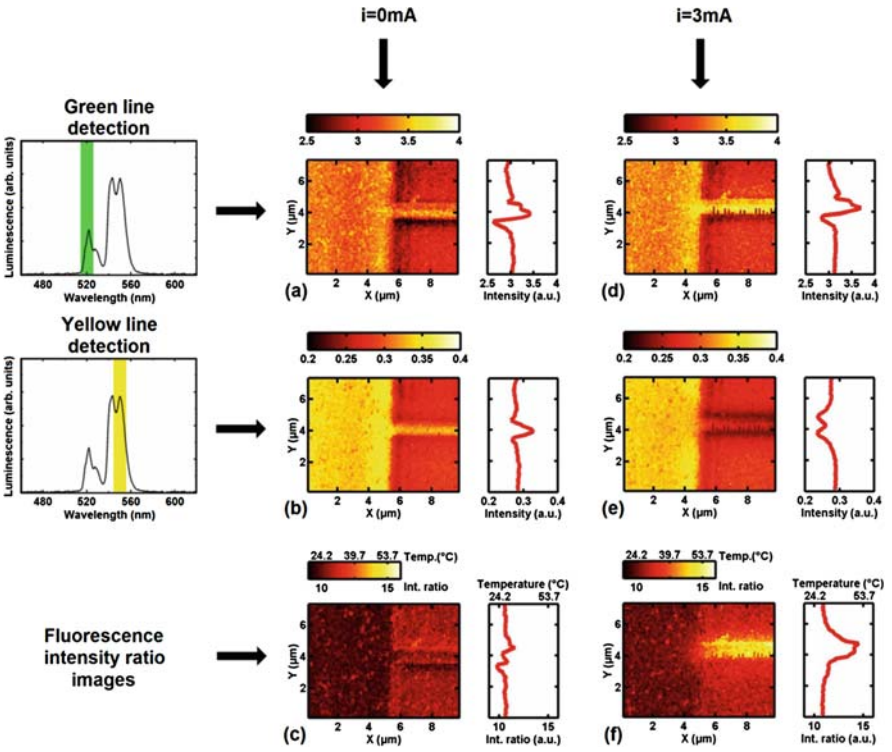

Fig. 17.16 Luminescence images of the particle while scanning a nickel strip of width 500 nm. (**a**), (**b**), and (**c**) are images obtained for $i = 0$ mA: (**a**) line at 520 nm, (**b**) line at 550 nm, (**c**) ratio I_{520}/I_{550} for the two luminescence lines. (**d**), (**e**), and (**f**) are the same as (**a**), (**b**), and (**c**), but for a current of 3 mA in the strip

only 50 nm of SiO. The role of the substrate is important in heat dissipation mechanisms. If the insulating layer is very thin, as it was for the aluminium sample, heat can be more or less directly evacuated into the substrate. On the other hand, if the insulating layer is thicker, as for the nickel strip, heat is less easily transmitted to the silicon and tends to diffuse laterally in the oxide layer. This explains why a slightly more spread-out heat spot is observed in the case of the nickel structure than in the case of the aluminium structure.

In order to assess the role of the oxide thickness in heat evacuation, it would be interesting to study a series of structures in which the thickness varies from a few nanometers to several tens of microns, and to carry out the deposits on substrates with higher thermal conductivity than silicon. Regarding the lateral resolution with this technique, the ideal test nanostructure is still the nanowire, e.g., carbon nanotubes, for which a lateral thermal resolution below 100 nm has been observed by Shi and Majumdar [40].

17.4.2 Alternating Current Measurements

The experiments discussed in the last section were carried out by passing a direct current through the device. Measurements can also be carried out when alternating current is passed through the structure. Indeed, in structures like metal nanowires, heating occurs very quickly and the material can reach its temperature after a few microseconds and cool down just as quickly. The aim of using current modulation is to increase the signal-to-noise ratio of the fluorescence measurement.

An Example of Thermal Images Obtained Using Alternating Current

In order to demonstrate the feasibility of such experiments, we passed an alternating current through a nickel strip. The temperature of the metal is then modulated by the Joule effect, inducing a modulation in the fluorescence. The basic idea is shown schematically in Fig. 17.17.

The device is excited by a current in the form of a positive periodic square function in such a way that the temperature is modulated at the same frequency as the current. The modulation frequency lies between 300 and 700 Hz depending on the experiment. The pump laser is not modulated and hence emits at constant power. In contrast to the direct current experiments, we measure here the variations in a single fluorescence line which, in the case of compounds doped with Er^{3+} ions, is the yellow–green line at 550 nm. Indeed, this line is the most sensitive to temperature. The experimental setup is a little simpler than for direct current measurements, because a single photomultiplier is sufficient [41]. We also use a slightly less selective interference filter in order to detect the whole luminescence line at 550 nm.

Figure 17.18 shows topographic and optical images obtained by scanning nickel strips of widths 4 μm and 200 nm deposited on a silicon oxide substrate. An increase in the optical signal is clearly visible when the particle is situated on the nickel strip. The measured signal is nevertheless a little more complex than for a continuous (direct current) excitation, because we have here a thermally induced fluorescence

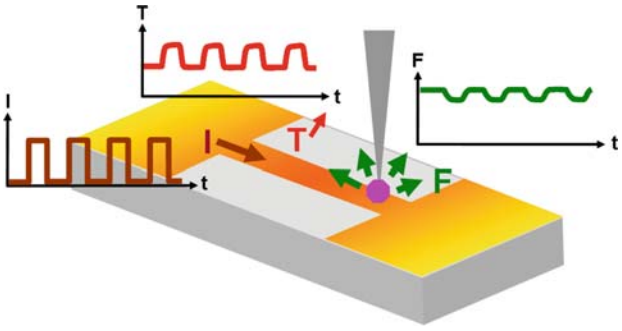

Fig. 17.17 Schematic view of thermal imaging experiments with alternating current

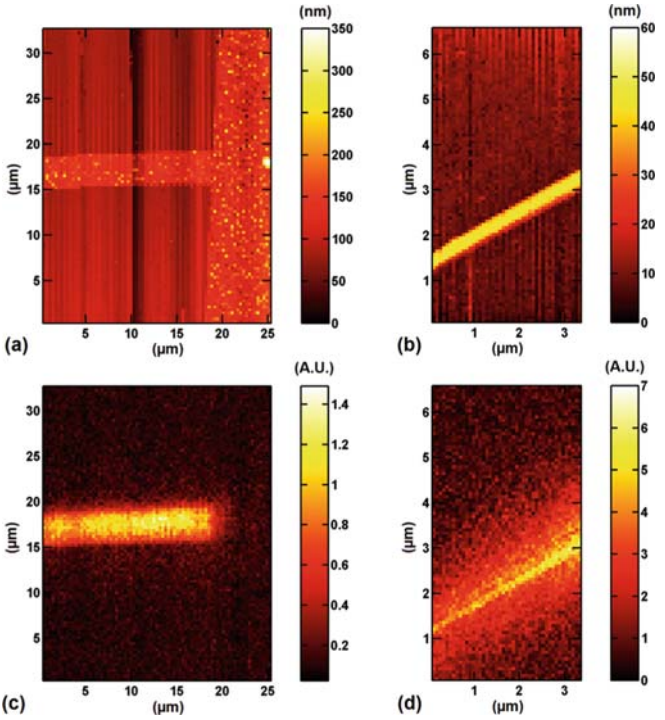

Fig. 17.18 (**a**) and (**b**) are topographic images and (**c**) and (**d**) thermally modulated fluorescence images obtained on nickel strips of widths 4 µm and 200 nm. The image on the 4 µm strip was taken at the join between strip and connector

gradient. If the particle is located at a point where the temperature does not vary, its emission is not modulated and the detected signal is zero. On the other hand, if it is on a region where the temperature varies, the fluorescence is modulated and a signal is detected.

Some care is needed when interpreting images obtained in alternating mode. It should always be borne in mind that the particle emits light whatever its position on the surface, and that the emitted intensity varies both with the temperature and with the intensity of the local field it is 'reading'. The signal S measured at the point with coordinates (x,y) on the surface can be written in a simplified way as the difference between the fluorescence intensity at (x,y) at the maximal temperature $T_{\max}$ and the same at the minimal temperature $T_{\min}$, i.e.,

$$S(x,y) = F^{T_{\max}}(x,y) - F^{T_{\min}}(x,y)\ . \tag{17.9}$$

Note that $T_{\max}$ and $T_{\min}$ also depend on the position (x,y), since the temperature is not uniform across the device. To get a more explicit expression for the measured signal, the contributions due to the temperature and local field distribution can be separated. We thus write the fluorescence as the product of two functions F_{map}

and f. The function F_{map} (assumed independent of temperature) represents the fluorescence distribution at the point (x,y) for a given uniform temperature across the surface, while the function f describes how the temperature affects light emission at the point (x,y):

$$F^{T_{\max}}(x,y) = F_{\text{map}}(x,y) f\left[T^{\max}(x,y)\right] , \quad F^{T_{\min}}(x,y) = F_{\text{map}}(x,y) f\left[T^{\min}(x,y)\right] . \tag{17.10}$$

The function f is different for each type of luminescent material. For example, if the fluorescence intensity of a material varies as $A/(B+T)$, where A and B are constants, then $f(x,y) = A/\left[B+T(x,y)\right]$. For the materials we used, the function f has a similar form to (17.1), and describes the depopulation mechanisms affecting the relevant energy level, viz., the level $^4S_{3/2}$ at 550 nm, by direct transfer to the neighbouring level $^2H_{11/2}$ at 520 nm, but also by other non-radiative mechanisms.

In order to obtain a description of the temperature distribution across the surface, or at least to approximate it, the measured signal $S(x,y)$ can be divided by the fluorescence distribution across the surface $F_{\text{map}}(x,y)$. We obtain

$$\frac{S(x,y)}{F_{\text{map}}(x,y)} = f\left[T^{\max}(x,y)\right] - f\left[T^{\min}(x,y)\right] . \tag{17.11}$$

This normalisation thus removes spatial variations in the fluorescence that are not related to temperature changes. The right-hand side of (17.11) is related to the temperature gradient at the point (x,y) of the sample surface.

Example of Normalisation in Alternating Mode

Figure 17.19 shows the topography of a nickel strip of width 200 nm, the thermally modulated fluorescence image $S(x,y)$, and the associated fluorescence image $F_{\text{map}}(x,y)$, measured at room temperature [41]. The latter shows that the particle emits more light on the nickel strip than on the substrate. Moreover, oscillations are visible on one side of the strip. These probably correspond to interference between the incident light and light diffracted by the nickel strip. Similar images are often observed with near-field optical microscopy [42].

As the images were obtained consecutively, the thermally modulated fluorescence image cannot be divided directly by the fluorescence image because, owing to thermal drift between the tip holder and the sample holder, they cannot be superposed. However, the profiles shown in Fig. 17.20, extracted from the two images and averaged over a dozen or so adjacent columns, can be fitted using the topographic profiles, and normalisation is then possible. The profiles represent the thermally modulated fluorescence $S(y)$, the room temperature fluorescence $F_{\text{map}}(y)$, and the ratio of the two.

The scales of the profiles $S(y)$ and $F_{\text{map}}(y)$ are different and there is no correlation between them. The slight shoulder visible at the edge of the strip on the $S(y)$ profile probably corresponds to the increase in fluorescence visible on the $F_{\text{map}}(y)$

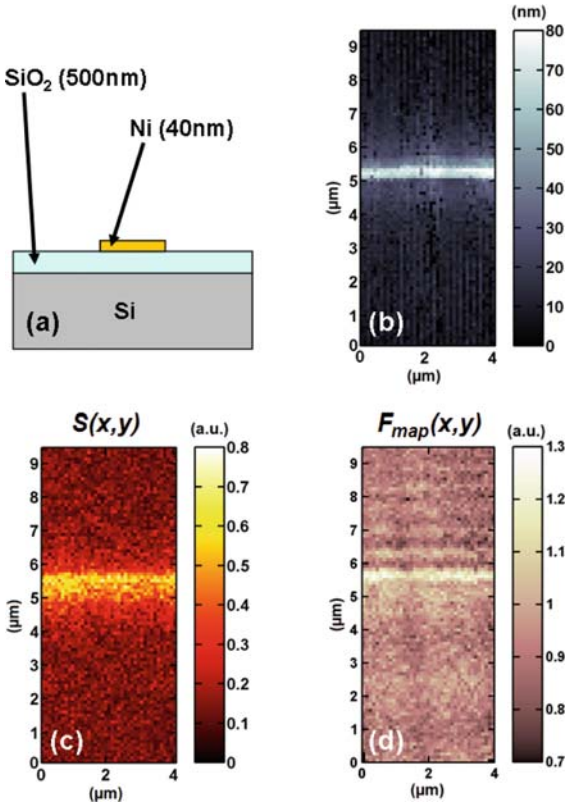

Fig. 17.19 (**a**) Schematic cross-section of the nickel strip. Thicknesses of the various materials are indicated. The width of the strip is 200 nm. (**b**) Topography. (**c**) Thermally modulated fluorescence image. (**d**) Fluorescence image obtained at room temperature

profile, and is not therefore of thermal origin. This shoulder is attenuated on the normalised profile $S(y)/F_{\mathrm{map}}(y)$. Likewise, the amplitude of the interference drops slightly after taking the ratio of the profiles. However, the effect of normalisation is not particularly great in the case of this structure, because the fluorescence variations are relatively small, not exceeding 20% of the average value over the whole image. It may happen in the case of samples carrying larger strips, or in the case of metal strips that reflect or scatter much more light such as gold or platinum, that these variations will be greater, whence the normalisation will have a significant effect on the fluorescence distribution.

Limits of the Alternating Mode and Future Improvements

Current modulation images remain difficult to analyse, even after normalising by the fluorescence distribution across the surface. Indeed, the normalised image represents

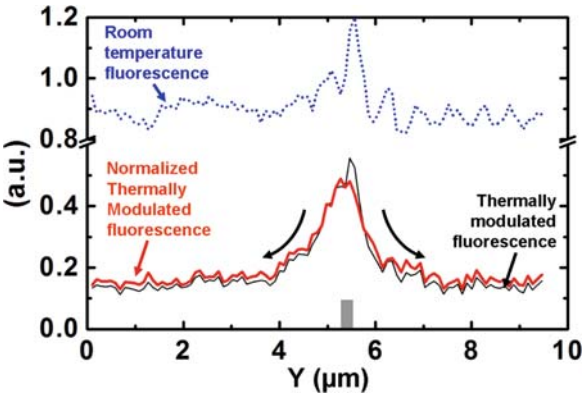

Fig. 17.20 Profiles extracted from the images in Fig. 17.19. The *thin black curve* and the *dotted blue curve* represent $S(y)$ and $F_{\text{map}}(y)$, respectively, while the *red curve* represents the ratio of these two profiles

the quantity $f\left[T_{\max}(x,y)\right] - f\left[T_{\min}(x,y)\right]$. There are then several ways to determine the real temperature at the surface. To begin with, we can determine the function f and extract $T_{\max}(x,y) - T_{\min}(x,y)$. f can be predicted either from a description of the relevant physical effects (see, for example, Sect. 17.2), or from a fit to experimental values.

Another effect that we have not taken into account is any delay that may occur between temperature modulation and fluorescence modulation. We have assumed that the two quantities are modulated in phase and that this phase does not vary while the sample is being scanned. However, this may not be the case and, over large distances, heat transmission between strip and particle may not be instantaneous, whereupon phase variations may arise. It would be very interesting to investigate such variations, for the results would provide a great deal of information concerning the propagation of heat in these structures.

The experiments were carried out at frequencies below the kilohertz. The current can be modulated at higher frequencies, but that would depend on the type of fluorescent particle used. In compounds doped with Er^{3+} ions, the fluorescence line we use has a lifetime between 100 μs and 1 ms, and this plays the role of a low-pass filter that limits its use. For applications requiring higher frequency studies, particles with faster radiative transitions would be necessary.

17.4.3 Measuring Tip–Sample Heat Transfer

We end this chapter by describing an experimental study of heat transfer between the tip and the sample. The temperature we measure is that of the fluorescent particle, which may differ from the sample temperature, because heat losses may occur

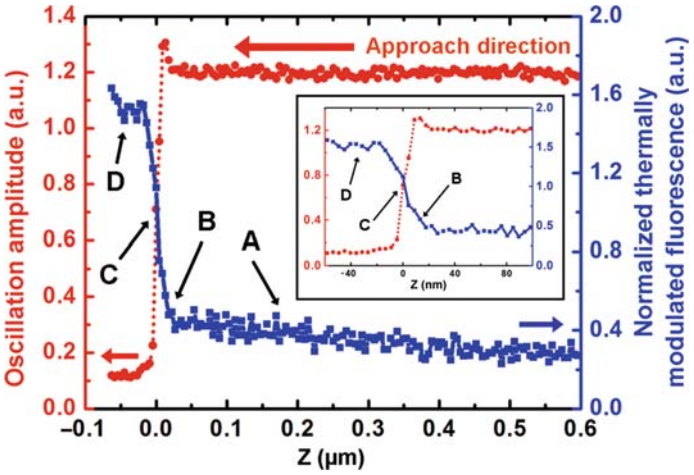

Fig. 17.21 Tip–surface approach curve. Oscillation amplitude (*red*) and thermally modulated fluorescence (*blue*) as a function of the distance between tip and sample. The *insert* shows the final approach more closely

in the tip–sample contact area. Approach curves provide a way understanding the mechanisms involved in heat transfer.

Several experimental [43, 44] and theoretical studies [45] have already been carried out using SThM. For our part, we have produced approach/retract curves over a heated strip. The experiment was carried out in the alternating current mode on a nickel strip of width 500 nm. The thermally modulated fluorescence curves and the vibration amplitude of the tip are shown in Fig. 17.21. At large distances from the strip (zone A of the curve), the fluorescence signal is not zero, indicating heat transfer by conduction through the air between strip and particle. This signal gradually increases until the two objects come into contact. At a distance of a few nanometers from contact, a slight increase in vibration amplitude is visible (zone B). This increase is very likely due to attractive forces between tip and sample. In this zone, the optical signal increases more significantly. Then the vibration amplitude gradually drops off until the tip is totally immobilised and the fluorescence increases considerably (zone C). The heat transfer mechanisms in this zone are more complex, because the tip alternates between being in contact with the strip and being free in the air. Heat transfer then results from a superposition of several phenomena: heating by direct contact and cooling during detachment. The more the oscillation amplitude is attenuated, the more time the tip spends in contact with the surface, and the more efficiently heat is transferred. Finally, when the tip has ceased to vibrate (zone D), the fluorescence signal stabilises at a certain value.

These curves suggest the following observations. To begin with, the images presented in the previous sections were obtained in tapping mode, i.e., with intermittent contact. During scanning, the oscillation amplitude is attenuated by about 50% as compared with free oscillation in the air. It is clear that heat transfer is not optimal

and that the measured temperature is less than the actual temperature of the strip. It should thus be borne in mind that, while the images give some idea of the temperature distribution over the device, the real temperature remains unknown. In order to measure that, the tip can nevertheless be placed at a precise point on the strip and immobilised there, and a signal measured under permanent contact conditions. This procedure is easy to put into practice.

Another point is that heat transfer through the air remains low, as the approach curve indicates. Naturally, this will depend on the size of the device under investigation. The bigger the heat source, the more efficient will be the heating by conduction through the air.

17.5 Conclusion and Prospects

We have shown that a fluorescent particle fixed to the end of an atomic force microscope tip can be used as a very efficient scanning thermal nanoprobe for characterising submicron devices. Many improvements can be made to the device we have described:

- Regarding improvement of the temperature measurement, it would obviously be preferable to work in permanent contact mode. This requires a different setup for measuring the tip deflection, either an optical technique using a laser, or some non-optical technique such as a piezoresistive cantilever.
- The lateral resolution of the probe can be improved by using smaller particles. One can envisage the use of particles of diameter less than 50 nm, if the aim is to characterise samples requiring thermal resolution of this order.
- Fast measurements can also be envisaged, provided one uses fluorescent particles with a shorter radiative lifetime than the ions we have used so far.
- Finally, other SThM techniques using a thermocouple [40,46] or thermoresistive wire [47, 48] require an electric current to carry out the temperature measurement. In the case of fluorescent probes, no current is needed. One could therefore envisage using the probe in a liquid medium for applications in biochemistry or microfluidics.

References

1. J.R. Lakowicz: *Principles of Fluorescence Spectroscopy*, Plenum, New York, 1999
2. Y.P. Varshni: Physica **34**, 149–154 (1967)
3. L. Viña, S. Logothetidis, and M. Cardona: Phys. Rev. B **30** (4), 1979–1991 (1984)
4. For a review, see for instance S.W. Allison and G.T. Gillies: Rev. Sci. Intrum. **68** (7), 2615–2650 (1997)
5. D. Ross, M. Gaitan, and L.E. Locascio: Anal. Chem. **73**, 4117–4123 (2001)
6. H.F. Arata, P. Löw, K. Ishizuka, C. Bergaud, B. Kim, H. Noji, and H. Fujita: Sensors & Actuators B **117**, 339–345 (2006)

7. F.H. Ko, L.Y. Weng, C.J. Ko, and T.C. Chu: Microelectron. Eng. **83**, 864–868 (2006)
8. M. Dejneka, E. Snitzer, and R.E. Riman: J. Lum. **65**, 227–245 (1995)
9. A.J. Kenyon: Prog. Quantum. Electron. **26**, 225–284 (2002)
10. F. Auzel: Chem. Rev. **104**, 139–173 (2004)
11. H. Berthou and C.K. Jörgensen: Opt. Lett. **15** (19), 1100–1102 (1990)
12. E. Maurice, G. Monnom, B. Dussardier, A. Saïssy, D.B. Ostrowsky, and G.W. Baxter: Appl. Opt. **34** (34), 8019–8025 (1995)
13. G. Paez and M. Strojnik: Appl. Opt. **42** (16), 3251–3258 (2003)
14. M.A.R.C. Alencar, G.S. Maciel, C.B. de Araujo, and A. Patra: Appl. Phys. Lett. **84** (23), 4753–4755 (2004)
15. W.H. Fonger and C.W. Struck: J. Chem. Phys. **52** (12), 6364–6372 (1970)
16. R.H. Krauss, R.G. Hellier, and J.C. McDaniel: Appl. Opt. **33** (18), 3901–3904 (1994)
17. H. Peng, H. Song, B. Chen, J. Wang, S. Lu, X. Kong, and J. Zhang: J. Chem. Phys. **118** (7), 3277–3282 (2003)
18. M.L. Bhaumik: J. Chem. Phys. **40** (12), 3711–3715 (1964)
19. S.M. Borisov and O.S.Wolfbeis: Anal. Chem. **78**, 5094–5101 (2006)
20. G. Laufer, N.B. Rotchford, and R.H. Krauss: Am. J. Phys. **65** (5), 447–449 (1997)
21. P. Kolodner and A.J. Tyson: Appl. Phys. Lett. **40** (9), 782–784 (1982)
22. K.T.V. Grattan and A.W. Palmer: Rev. Sci. Instrum. **56** (9), 1784–1787 (1985)
23. S.W. Allison, G.T. Gillies, A.J. Rondinone, and M.R. Cates: Nanotechnology **14**, 859–863 (2003)
24. Z. Zhang, K.T.V. Grattan, and A.W. Palmer: Phys. Rev. B **48** (11), 7772–7778 (1993)
25. M.A. Hines and P. Guyot-Sionnest: J. Phys. Chem. **100**, 468–471 (1996)
26. B.O. Dabbousi, J. Rodriguez-Viejo, F.V. Mikulek, J.R. Heine, H. Matoussi, R. Ober, K.F. Jensen, and M.G. Bawendi: J. Phys. Chem. B **101**, 9463–9475 (1997)
27. G.W. Walker, V.C. Sundar, C.M. Rudzinski, A.W. Wun, M.G. Bawendi, and D.G. Nocera: Appl. Phys. Lett. **83** (17), 3555–3557 (2003)
28. D. Valerini, A. Creti, M. Lomascolo, L. Manna, R. Cingolani, and M. Anni: Phys. Rev. B **71**, 235409 (2005)
29. V. Biju, Y. Makita, A. Sonoda, H. Yokoyama, Y. Baba, and M. Ishikawa: J. Phys. Chem. **109**, 13899–13905 (2005)
30. A. Joshi, K.Y. Narsingi, M.O. Manasreh, E.A. Davis, and B.D. Weaver: Appl. Phys. Lett. **89**, 131907 (2006)
31. A.J. Melmed: J. Vac. Sci. Technol. B **9** (2), 601–608 (1991)
32. L. Aigouy, Y. De Wilde, M. Mortier, J. Giérak, and E. Bourhis: Appl. Opt. **43** (19), 3829–3837 (2004)
33. See, for instance, L. Wang: Appl. Phys. Lett. **73** (25), 3781–3783 (1998)
34. P. Kolodner and J.A. Tyson: Appl. Phys. Lett. **42** (1), 117–119 (1983)
35. E. Van Keuren, M. Cheng, O. Albertini, C. Luo, J. Currie, and M. Paranjape: Sensors and Materials **17** (1), 1–6 (2005)
36. P. Löw, B.J. Kim, N. Takama, and C. Bergaud: Small **4** (7), 908–914 (2008)
37. L. Aigouy, G. Tessier, M. Mortier, and B. Charlot: Appl. Phys. Lett. **87**, 184105 (2005)
38. B. Samson, L. Aigouy, R. Latempa, G. Tessier, M. Aprili, M. Mortier, J. Lesueur, and D. Fournier: J. Appl. Phys. **102**, 024305 (2007)
39. B. Samson, L. Aigouy, G. Tessier, P. Löw, C. Bergaud, B. Kim, and M. Mortier: J. of Phys.: Conference Series **92**, 012089 (2007)
40. L. Shi, S. Plyasunov, A. Bachtold, P.L. McEuen, and A. Majumdar: Appl. Phys. Lett. **77** (26), 4295–4297 (2000)
41. B. Samson, L. Aigouy, P. Löw, C. Bergaud, B. Kim, and M. Mortier: Appl. Phys. Lett. **92**, 023101 (2008)
42. J.J. Greffet and R. Carminati: Prog. Surf. Sci. **56**, 133–237 (1997)
43. S. Gomès, N. Trannoy, and P. Grossel: Meas. Sci. Technol. **10**, 805–811 (1999)
44. L. Shi and A. Majundar: J. Heat Transfer **124**, 329–337 (2002)
45. P.O. Chapuis, J.J. Greffet, K. Joulain, and S. Volz: Nanotechnology **17**, 2978–2981 (2006)

46. G. Mills, H. Zhou, A. Midha, L. Donaldson, and J.M.R. Weaver: Appl. Phys. Lett. **72** (22), 2900–2902 (1998)
47. H.M. Pollock and A. Hammiche: J. Phys. D Appl. Phys. **34**, R23–R53 (2001)
48. S. Lefèvre and S. Volz: Rev. Sci. Instrum. **76**, 033701 (2005)

Chapter 18
Heat Transfer in Low Temperature Micro- and Nanosystems

Olivier Bourgeois

The study of thermal and thermodynamic properties at the nanoscale requires the development of samples with well controlled small scale structure, but also ultrasensitive and innovative experimental techniques for handling such samples. The challenge is to measure very small amounts of energy, and to control the flow of these energies on very small length scales. Such measurements generally depend on very precise temperature control made possible by ultrasensitive thermometry. From this point of view, electrical measurements afford unique solutions, because they are easily adapted to small scales by exploiting experimental techniques developed to measure electrical resistances. With the help of technologies transferred from micro- and nanoelectronics, devices and sensors can be designed to measure the physical properties of small systems. In this chapter, we begin by calculating the thermodynamic properties expected for condensed matter at low temperatures. The temperature dependence of the specific heat and the thermal conductivity are calculated for each type of heat carrier, viz., phonons and electrons. Special attention is paid to the specificities of low-dimensional systems: quantum effects on the thermal conductance and the heat capacity. We then describe the experimental aspects (techniques and instrumentation), by reviewing the various solutions available in thermometry, and methods for measuring the specific heat and thermal conductivity, in either steady state or dynamical contexts. We will see how to apply each technique on the submicron scale, illustrating with different suspended systems in the case of membranes and nanowires.

18.1 Introduction

In this chapter, we will be concerned with the thermal properties of micro- and nanostructured matter at low temperatures. The physics of low-dimensional solid state matter is pushed to an extreme at low temperatures and looks very different from that of the corresponding bulk material. Several characteristic physical length

S. Volz (ed.), *Thermal Nanosystems and Nanomaterials*, Topics in Applied Physics, 118
DOI 10.1007/978-3-642-04258-4_18, © Springer-Verlag Berlin Heidelberg 2009

scales will play a determining role, some increasing when the temperature decreases and entering into competition with the sample dimensions. The main examples are:

- The mean free path Λ of the phonons, treating phonon transport as ballistic.
- The dominant wavelength λ_{dom} of the phonons.
- The mean free path l_{e} of the electrons.
- The quantum phase coherence length L_{φ} of the electron.

These physical quantities will determine the thermal and electrical behaviour of the cold matter. It is important to ascertain the laws governing their evolution, to obtain a better understanding of the physics involved and to be able to carry out controlled electrical measurement of thermal properties. We shall see that certain phenomena specific to temperatures close to absolute zero will perturb the measurements (mismatch of phonon velocities between two solids, decoupling of the phonon and electron thermal baths, etc.), and must be taken into account when interpreting experimental results. We shall focus on the thermal specificities of small systems, i.e., systems that are small compared with some characteristic physical length scale, which precludes reference to any particular size, although interesting dimensions are generally in the range between the nanometer and the micron.

Size reduction of physical systems has thermal consequences that we shall discuss below:

- Surface/volume thermodynamic competition (loss of bulk behaviour).
- Consequences of boundary conditions on artificial nanostructures.
- Effect of phonon confinement on the specific heat and thermal conductance.
- Importance of fluctuations in small systems.
- Existence of phase transitions specific to small length scales.

Many works treat this subject in the context of solid state physics [1], low temperature physics [2, 3], and the physics of electrons and phonons [4], while the specific heat is discussed in [5,6] and nanoscale energy transfer in [7]. The interested reader is referred to these for more information.

18.2 Thermal Physics at Low Temperatures

Before considering the thermal aspects of condensed matter at low temperatures and at small length scales, we begin by outlining the corresponding physics at the macroscopic scale. We shall discuss thermal properties, i.e., thermodynamics and transport, for the two main heat carriers in solids, viz., elastic waves of the crystal lattice (phonons) and electrons.

18.2.1 Equilibrium Thermodynamics

Specific Heat of Phonons

The microscopic description of specific heat, which is defined as the capacity of a body to rise in temperature following the supply of heat from the outside, was only achieved at the beginning of the twentieth century. The first theory is due to Einstein in 1907, which describes solids as ensembles of independent atoms that vibrate at a well defined frequency. This theory gives good results at high temperatures (as if one restricted to optical modes), but provides a poor description for observations at low temperatures. It was not until 1912 that Debye gave a more realistic description by treating elastic vibrations as a continuum with a cutoff frequency corresponding to the interatomic distance. It was only with the advent of this theory that physicists could account for the T^3 temperature dependence of the specific heat of electrically insulating materials at low temperatures. We shall outline the main features below.

Three-Dimensional Systems at Low Temperature

The specific heat is given by the variation of the internal energy as a function of temperature, and is thus expressed as $C_\mathrm{v} = \mathrm{d}U/\mathrm{d}T$. We begin by calculating the internal energy of the solid, assuming that it is a sum of quantum harmonic oscillators satisfying Bose–Einstein statistics:

$$U = \sum_k \left[\frac{1}{2}\hbar\omega(\mathbf{k}) + \frac{\hbar\omega(\mathbf{k})}{\mathrm{e}^{\hbar\omega(\mathbf{k})/k_\mathrm{B}T} - 1} \right] . \tag{18.1}$$

Rewriting this sum as an integral using the continuum hypothesis, the internal energy becomes

$$U = \int_\mathrm{BZ} \frac{\hbar\omega(\mathbf{k})}{\mathrm{e}^{\hbar\omega(\mathbf{k})/k_\mathrm{B}T} - 1} \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3} = \int \frac{\hbar\omega}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T} - 1} D(\omega)\mathrm{d}\omega . \tag{18.2}$$

We drop the first term in (18.1), which corresponds to what is known as the zero point energy. It has no temperature dependence and so plays no role in the specific heat.

For a first calculation at low temperature, we shall assume that only the long wavelengths, of low energy since $k_\mathrm{B}T$ is small, are excited. Then with only a small error, the integral to be carried out over the first Brillouin zone (BZ) can be extended over the whole of k space in three dimensions ($\mathrm{d}\mathbf{k} = 4\pi k^2 \mathrm{d}k$, infinitesimal volume element).

We shall make some further assumptions. First, we consider an average speed of sound v_s for the three vibration modes: three acoustic branches, two transverse and one longitudinal (the optical modes are not populated because they correspond to much higher energies compared with $k_\mathrm{B}T$). We shall also assume a dispersion

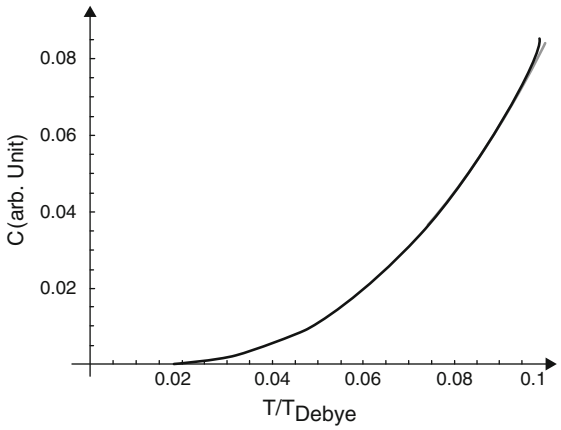

Fig. 18.1 Normalised temperature dependence of the specific heat at low temperatures, illustrating the T^3 variation

relation of the form $\omega(k) = v_s k$, which is justified at low temperatures. Then

$$U = \frac{6}{\pi^2} \int_0^\infty \frac{\hbar v_s k^3 \mathrm{d}k}{\mathrm{e}^{\hbar v_s k / k_B T} - 1} \,. \tag{18.3}$$

With a suitable change of variables, this becomes

$$U = \frac{3}{2\pi^2} \frac{k_B^4 T^4}{\hbar^3 v_s^3} \int_0^\infty \frac{x^3 \mathrm{d}x}{\mathrm{e}^x - 1} \,, \tag{18.4}$$

whence the specific heat is given by

$$C_{\mathrm{ph}} = \frac{2\pi^2}{5} \frac{k_B^4 T^3}{\hbar^3 v_s^3} \,. \tag{18.5}$$

Making appropriate assumptions for low temperature solid state physics, we can thus show that, in three dimensions, the specific heat varies as the cube of the temperature, in perfect agreement with experiment (see Fig. 18.1).

Debye Model

We shall now describe the Debye model, which provides the link with higher temperatures. The hypotheses are almost the same, namely, a linear dispersion relation and a continuum hypothesis. However, the integral over k space will now have to include a high frequency cutoff, which is a consequence of the discrete atomic structure. Indeed, at higher temperatures, higher energy levels will be excited, but one cannot excite modes whose wavelengths would be shorter than the interatomic distance. This limit implies a maximal wave vector (the Debye wave vector) $k_D^{\max} = 2\pi/\lambda_D^{\max}$,

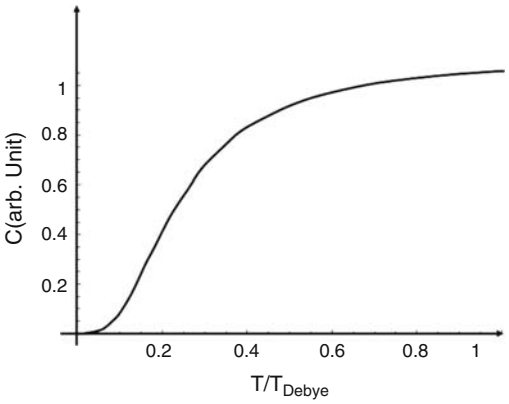

Fig. 18.2 Specific heat as a function of normalised temperature in the Debye model

where $\lambda_{\mathrm{D}}^{\max}$ is the interatomic distance. This wave vector is then used to define the Debye temperature θ_{D} by $k_{\mathrm{B}}\theta_{\mathrm{D}} = \hbar v_{\mathrm{s}} k_{\mathrm{D}}^{\max}$. Above this temperature, we can say that all possible modes in the crystal will be excited. Naturally, these hypotheses will only be valid if $\lambda_{\mathrm{D}}^{\max}$ is very small compared with the dimensions of the given solid. Equation (18.3) then becomes

$$U = \frac{6}{\pi^2} \int_0^{k_{\mathrm{D}}^{\max}} \frac{\hbar v_{\mathrm{s}} k^3 \mathrm{d}k}{\mathrm{e}^{\hbar v_{\mathrm{s}} k / k_{\mathrm{B}} T - 1}} \,. \tag{18.6}$$

With a suitable change of variable, we will thus find for the specific heat at any temperature

$$C_{\mathrm{ph}} = \frac{3}{2\pi^2} \frac{k_{\mathrm{B}}^4 T^3}{\hbar^3 v_{\mathrm{s}}^3} \int_0^{\theta_{\mathrm{D}}/T} \frac{\mathrm{e}^x x^4 \mathrm{d}x}{(\mathrm{e}^x - 1)^2} \,. \tag{18.7}$$

The temperature dependence of the specific heat is plotted in Fig. 18.2. The low temperature limit of (18.7) recovers the result obtained in (18.5), i.e., a T^3 dependence for $C_{\mathrm{ph}}(T)$.

Specific Heat of Low-Dimensional Systems

In fact, we have just shown that, in the Debye model, the specific heat is proportional to T^d, where d is the dimension of the object under consideration. To find out whether a plane system or wire can be considered to have lower dimension than a 3D object, one must compare the shortest length scales with the relevant characteristic length, which is here the dominant phonon wavelength λ_{dom}. This wavelength corresponds to the maximum spectral density of phonon black body radiation obeying Planck's law. It is thus given by $\lambda_{\mathrm{dom}} = hv_{\mathrm{s}}/2.82k_{\mathrm{B}}T$ [8].

This means that, for a planar system such as a membrane, when the transverse dimension becomes greater than λ_{dom} (at low enough temperature), the system can

be treated as a 2D phonon gas, and its specific heat will be quadratic in the temperature [9]. The phonon spectrum in the perpendicular direction can no longer be considered as a continuum, i.e., it thus becomes discrete. This T^2 behaviour had already been observed a long time ago on highly 2D compounds like graphite, where the different layers of the material are strongly decoupled [4, 10, 11].

For 1D systems, the reasoning is the same. We expect the specific heat to be linear in the temperature. For systems like carbon nanotubes, several experiments have confirmed these predictions. A linear specific heat has been measured on multiwalled carbon nanotubes by a Chinese group [12], up to very high temperatures (100 K). This would be the signature of a very high Debye temperature, which is expected for carbon nanotubes ($T_{\text{Debye}} \approx 2\,300$ K). Measurements on bundles of single-walled carbon nanotubes confirm the linear T dependence and demonstrate the presence of a contribution going as T^{2-3} at higher temperatures, due to coupling between the nanotubes, thereby restoring the 3D behaviour.

This discussion is only valid if the dispersion relation is linear, which is true for many materials. For greater generality, if $\omega \propto k^{\delta}$, then the specific heat will vary as $T^{d/\delta}$. For example, this is what would be expected for graphene. Although it is 2D, it must have $C_{\text{ph}} \propto T$, because the dispersion relation is quadratic [13, 14].

Specific Heat of Electrons

The increase in temperature of a body following the supply of heat can occur through excitation of degrees of freedom associated with the crystal lattice, as we have just seen with the phonons, but also through other degrees of freedom relating to the electrons. The specific heat of electrons is found in the same way as the specific heat of phonons. We begin by calculating the total energy and examining its temperature dependence.

Noble Metals

In the simplest case, the electrons in a metal can be treated as a free electron gas. This gas of indistinguishable particles satisfies Fermi–Dirac statistics as a consequence of the Pauli exclusion principle. One then speaks of a gas of uncorrelated fermions. In three dimensions, these objects can thus be treated as plane waves with kinetic energy $E = \hbar^2 k^2 / 2m$.

The total energy of the electrons is then

$$U = \int_0^{\infty} E f(E) D(E) \mathrm{d}E \,, \tag{18.8}$$

where

$$f(E) = \frac{1}{1 + \exp\left[(E - E_{\text{F}})/k_{\text{B}}T\right]}$$

is the Fermi–Dirac distribution and $D(E)$ the electron density of states. The latter is given by

$$D(E) = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{E}\,.$$

The Fermi energy E_F is defined as the maximum energy level of electrons in a solid at zero temperature. For a good metal, E_F is of the order of a few eV.

At this stage, the integral giving the specific heat directly is very complicated. However, we may make certain very useful assumptions from the fact that the Fermi temperature defined by $k_B T_F = E_F$ corresponds to a few times 10^4 K, so that even at room temperature we may assume that $k_B T \ll E_F$. Consequently, the energy levels of excited electrons will always remain very close to E_F, and the Fermi–Dirac function will only therefore vary appreciably close to the Fermi energy. We shall thus make two assumptions: $k_B T \ll E_F$ and $D(E) = D(E_F)$, i.e., the density of states will be taken as constant in the region where $\mathrm{d}f(E)/\mathrm{d}T$ is varying.

The specific heat will then be given by

$$C_{e^-} = \int_0^\infty (E - E_F)\frac{\mathrm{d}f(E)}{\mathrm{d}T}D(E)\mathrm{d}E = k_B^2 T D(E_F)\int_{-\infty}^{\infty}\frac{e^x x^2 \mathrm{d}x}{(e^x+1)^2}\,, \tag{18.9}$$

whence we obtain a linear variation of C_{e^-} with temperature:

$$C_{e^-} = \frac{\pi^2}{3}D(E_F)k_B^2 T\,. \tag{18.10}$$

At room temperature, it is the phonon specific heat that dominates, whereas at low enough temperatures, it is the linear term in the electron specific heat in a metal that will become predominant.

Two-Level System

For a system with two discrete energy levels (spin 1/2 electrons in a magnetic field), it is useful to calculate the specific heat at low temperature, because there is a large anomaly that can significantly perturb measurement results. Take the case of a physical system with two accessible levels ε_1 and ε_2. Using statistical physics, we can calculate the average energy of a two-level system at temperature T [15]. We begin by writing down the partition function for this system, from which we deduce its energy:

$$Z = \sum_{i=1}^{2} e^{-\varepsilon_i/k_B T}\,, \qquad E = \frac{1}{Z}\sum_{i=1}^{2}\varepsilon_i e^{-\varepsilon_i/k_B T} = \varepsilon_1 + \delta\frac{e^{-\delta/k_B T}}{1+e^{-\delta/k_B T}}\,, \tag{18.11}$$

where $\delta = \varepsilon_2 - \varepsilon_1$ is the difference between the energy levels. The specific heat of the two-level system will thus be given by $C_{2l} = \mathrm{d}E/\mathrm{d}T$:

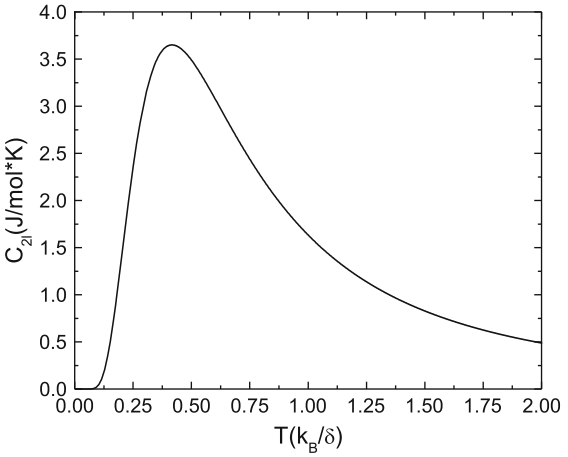

Fig. 18.3 Temperature dependence of C_{2l}. The Schottky anomaly is centered on the temperature $T = 0.42k_{\mathrm{B}}/\delta$

$$C_{2l} = \frac{\delta^2}{k_{\mathrm{B}}T^2} \frac{\mathrm{e}^{-\delta/k_{\mathrm{B}}T}}{(1+\mathrm{e}^{-\delta/k_{\mathrm{B}}T})^2} \,. \tag{18.12}$$

This expression reflects an anomaly at low temperatures, called the Schottky anomaly, centered on the temperature $T = 0.42k_{\mathrm{B}}/\delta$. For a better view of this anomaly, the temperature dependence of C_{2l} is shown in Fig. 18.3.

18.2.2 Quasi-Steady State Nonequilibrium Heat Transfer

In this section, we shall calculate the energy transfer by phonons and electrons under steady-state conditions, without considering transient intermediate regimes. The problem here is thus to calculate the thermal conductance of various systems depending on the statistics of the energy carriers and the specific geometry of the conductors, taking into account the distribution of these particles in phase space $(\mathbf{r}, \mathbf{p})$.

Thermal Conductivity. Kinetic Method

The thermal conductance is a physical quantity depending on the kind of material, but also its topology. The thermal conductance is defined as the coefficient of proportionality between the thermal gradient set up when power is supplied from the outside and the amount of that power. In order to calculate this, we begin by expressing what we mean by the heat flux. Indeed, we know that if there is a temperature gradient along the x axis, a certain amount of heat will flow along this axis. Intuitively, this flux can be described by

$$\phi = \frac{1}{V}\sum_{k} v_x E(k) f(\mathbf{r},\mathbf{p}) \,, \tag{18.13}$$

where ϕ is the heat flux (in W/m^2), V the volume of the system, v_x the particle speed along the relevant axis, E the energy of the given particles, and f their distribution function in the phase space $(\mathbf{r},\mathbf{p})$. By using this kinetic formulation, we are sure to have accounted for all the particles with their corresponding energies and speeds and as a function of their distribution.

Going to the continuous limit, (18.13) can be rewritten in the form

$$\phi = \frac{1}{V}\sum_{s} \iiint \mathrm{d}\mathbf{k}\; v_x E(k) f(\mathbf{r},\mathbf{p}) \frac{L^3}{2\pi} \,, \tag{18.14}$$

transforming the integral over $\mathbf{k}$ to an integral over angular frequencies. The heart of the problem will be to calculate the exact value of this distribution function. The function f satisfies Boltzmann's equation, which treats the energy carrying particles as a dilute gas of classical particles. This equation expresses the evolution of the particle flow in phase space (Liouville equation):

$$\frac{\partial f}{\partial t} + \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\cdot\nabla f + \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}\cdot\nabla f = \left.\frac{\partial f}{\partial t}\right|_{\mathrm{coll}} \,. \tag{18.15}$$

The right-hand side of this equation is called the collision integral. It embodies all the scattering processes occurring in a real system. This term can be simplified using the relaxation time approximation:

$$\left.\frac{\partial f}{\partial t}\right|_{\mathrm{coll}} = -\frac{f - f_0}{\tau} \,, \tag{18.16}$$

where f_0 is the equilibrium distribution function and τ a characteristic scattering time. Using (18.16), we can solve (18.15) by making a simple steady state hypothesis, viz., $\partial f/\partial t = 0$, and assuming that there is no external force, so that $\mathrm{d}\mathbf{p}/\mathrm{d}t = 0$. The solution of (18.15) is then

$$f(r,k) = f_0 - \tau\frac{\mathrm{d}f}{\mathrm{d}T}\mathbf{v}\cdot\nabla T \,. \tag{18.17}$$

This equation shows that the presence of a temperature gradient in one space direction will create an imbalance between the flows of phonons or electrons. This imbalance will then lead to energy transfer. From this solution, we can calculate the integral given in (18.14) for the two different energy carriers we are interested in, namely, electrons and phonons, taking into account among other things their specific statistics, the densities of states, and scattering.

Thermal Conductivity and Conductance of Phonons

We begin by considering the case of a temperature gradient in the x direction. We can then rewrite (18.14) using the expression for the phonon energy, their density of states $D(\omega) = \mathrm{d}k/\mathrm{d}\omega$, and (18.16), whence we obtain the following expression for the heat flux:

$$\phi = \sum_s \int_0^{\omega_{\max}} \mathrm{d}\omega \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\theta \; v_x \hbar\omega \left(f_0 - \tau \frac{\mathrm{d}f}{\mathrm{d}T} \mathbf{v} \cdot \nabla T \right) \frac{D(\omega)}{4\pi} \,. \tag{18.18}$$

If we are only interested in a temperature gradient along the x axis, then $\nabla T = \mathrm{d}T/\mathrm{d}x$. We then directly obtain the Fourier law:

$$\phi = -k_{\mathrm{ph}} \frac{\mathrm{d}T}{\mathrm{d}x} \,. \tag{18.19}$$

The factor k_{ph} here is called the thermal conductivity, and it is given by

$$k_{\mathrm{ph}} = \sum_s \int_0^{\omega_{\max}} \mathrm{d}\omega \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\theta \; \sin\theta \cos^2\theta \, \hbar\omega \tau v_x^2 \frac{\mathrm{d}f_0}{\mathrm{d}T} \frac{D(\omega)}{4\pi} \,. \tag{18.20}$$

Returning to (18.2), we recover the expression for the specific heat already calculated. In this kinetic model, we may thus establish a direct relation between the thermal conductivity and the specific heat:

$$k_{\mathrm{ph}} = \frac{1}{3} C_{\mathrm{ph}} v_{\mathrm{ph}} \Lambda \,, \tag{18.21}$$

where the elastic mean free path is defined as $\Lambda = \tau v_{\mathrm{ph}}$, since τ as defined in the Boltzmann equation is related to scattering within the system. The last relation can be used to calculate the thermal conductivities from what we have already calculated in the last section on the specific heat. For example, at low temperatures, the thermal conductivity of a bulk material will go as T^3, like the specific heat.

Surface Effects

Up to now we have always considered our systems to be infinite. This condition is effectively fulfilled when the phonon mean free path is small compared with the characteristic dimensions of the system. The thermal conductivity is then a quantity independent of the system geometry. These assumptions fail when we go to small length scales. Indeed, in the case of a nanowire for example, the diameter is generally much smaller than the mean free path, and scattering at the edge of the conductor will enter into the calculation of the thermal conductance. This is particularly true at low temperatures, where the various scattering processes become more scarce. For infinitely rough surfaces, the mean free path will be given by the diameter D of the cylindrical nanowire, i.e., $\Lambda = D$ (Casimir theory) [4]. Equation

(18.21) can then be rewritten to obtain the thermal conductivity:

$$k_{\mathrm{ph}} = \frac{1}{3} C_{\mathrm{ph}} v_{\mathrm{ph}} D \, .$$

However, care must be taken here, because the very concept of thermal conductivity is inapplicable when its value depends on the system geometry. One should thus speak of thermal conductance rather than conductivity, taking into account the shape of the sample.

When we consider infinitely rough surfaces, we assume that all phonon wavelengths are scattered in the same way, i.e., that the surfaces behave as phonon black bodies with emissivity $\varepsilon = 1$. Now at low dimensions in particular, this will not be justified, because some phonons will be specularly reflected while others will be scattered. Surfaces must then be treated as imperfectly rough. Under such intermediate conditions, it is intuitively clear that short wavelengths ($\lambda < 1$ nm) will be mainly scattered, because asperities will be of the same order of magnitude, whereas long wavelengths, which dominate at low temperatures, will be specularly reflected. These specular reflections are not taken into account in the Boltzmann equation as we have expressed it, so appropriate models must be developed. The transition from rough to smooth for different temperatures will make the transport more and more ballistic, with important consequences, as we shall see below, especially for nanoscale objects [16, 17].

Interfacial Thermal Resistance

In many thermal micro- and nanosystems, very different materials are found together, e.g., thermometer, mechanical support, heating, and if possible, these must be mutually thermalised. However, the region separating any two materials will have an interfacial thermal resistance, often called the Kapitza resistance, which becomes particularly important at low temperatures. There are two models for evaluating this resistance, describing the two limits of the heat exchange process between the media: the acoustic mismatch model (AMM) and the diffuse mismatch model (DMM) [18].

The first of these, the AMM, assumes that no scattering phenomenon occurs at the surface and that the different speeds of sound in the two materials limit phonon exchange, giving rise to an interfacial thermal resistance. In contrast, the DMM assumes that any phonon striking the interface will be scattered, so that this scattering restores thermal conduction between the media. The quality of the interfaces will thus be a crucial element in the evaluation of this resistance. Whatever model is considered, the interfacial thermal conductance will be given by $k = \alpha T^3$ in W/K cm^2. The coefficient α will depend on the model and will be a function of the phonon speeds in each medium and the density of the materials. Apart from the T^3 variation, which explains the decrease in interfacial conductance at low temperatures ($T \leq 4$ K), we see that for very small surfaces these thermal resistances will be large and may perturb experiments on nanoscale objects by limiting the

thermalisation between the various elements of the system. As a consequence, they must be taken into account when setting up cold thermal experiments, e.g., between copper and liquid helium, we have $\alpha_{\mathrm{Cu/He}} = 2 \times 10^{-3}$ W/K^{4}cm^2, while in the case of solid–solid contacts, here copper–silicon, the value is more favourable, viz., $\alpha_{\mathrm{Cu/Si}} = 7 \times 10^{-2}$ W/K^{4}cm^2 (values taken from [18]).

Nanoscale Phonon Transfer

Over the last decade or so, theoreticians and experimenters have been interested in the limiting cases for energy and heat exchange between nanoscale conductors at low temperatures [19, 20]. This is not a new problem, since Pendry had already shown in the 1980s that information transfer, and also energy and entropy transfer, have a limiting value [21, 22]. We seek to describe the heat transfer between two reservoirs held at different temperatures and connected by a thermal conductor with very small dimensions (see Fig. 18.4).

The low temperature limit is interesting because it can greatly simplify the expression for the thermal conductance. Indeed, we shall assume that the phonons have very long mean free paths compared with the typical dimensions of the system ($\Lambda \gg L$), whence the transport between R_1 and R_2 will be ballistic. Moreover, if the diameter d of the conductor connecting the reservoirs is very small compared with the dominant phonon wavelength, the transport can be treated as one-dimensional ($\lambda_{\mathrm{dom}} > d$). From these hypotheses, and returning to the equation (18.13) for the heat flux due to a particle flow, we can express the total heat exchange between the two reservoirs at equilibrium, for $T_1 > T_2$, in the form

$$\Phi_{1\to 2} = \phi_1 - \phi_2 = \frac{1}{S}\sum_s \int \mathrm{d}k_x \, v_x \left[f_0(k, T_1) - f_0(k, T_2) \right] , \tag{18.22}$$

where S is the cross-sectional area of the 1D conductor and the transmission coefficients between R_1 and R_2 are equal to 1. If the temperature gradient is assumed to be small compared with the change in the distribution function, then (18.22) implies

$$K = \frac{\Phi_{1\to 2} S}{T_1 - T_2} = \sum_s \int_{\omega_{\min}}^{\omega_{\max}} \hbar\omega \frac{\mathrm{d}f}{\mathrm{d}T} \mathrm{d}\omega \, . \tag{18.23}$$

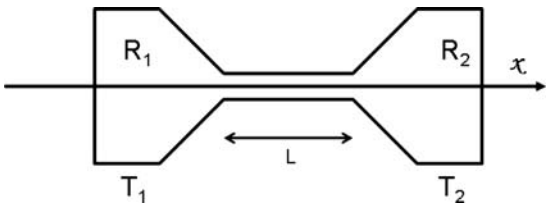

Fig. 18.4 Nanowire connected to two heat reservoirs, with heat flux along the x axis

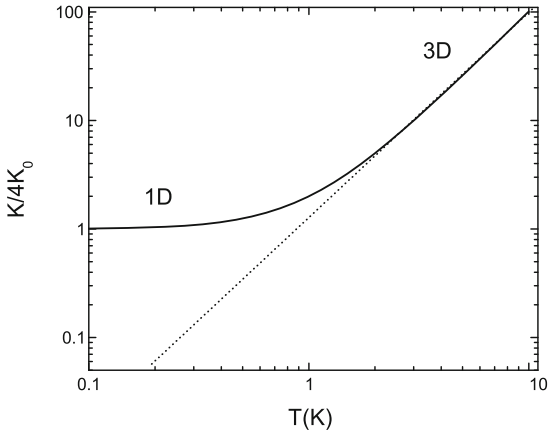

Fig. 18.5 Temperature dependence of the thermal conductance normalised at $K_0 = \pi^2 k_B^2 T/3h$. The curve illustrates the transition of the thermal conductance of a 1D conductor from 3D to 1D behaviour with falling temperature

Taking $\omega_{max} = \omega_{Debye} \approx \infty$ and $\omega_{min} \propto 1/L \approx 0$, and bearing in mind that there are four polarisations in a 1D conductor, this leads to the expression for the thermal conductance of a 1D conductor [24]:

$$K = 4\frac{\pi^2 k_B^2 T}{3h} \,. \tag{18.24}$$

So at low enough temperatures, the thermal conductance becomes independent of the material. It takes a universal value, linear in the temperature, equal to $K_0 = \pi^2 k_B^2 T/3h$ (because 1D) for each conduction channel (there being 4 here). This behaviour should be compared with the quantization of electrical conductance in a 1D system. In the case of thermal conductance, one cannot strictly speak of quantization, since one only observes a saturation of K when it is normalised by K_0 (see Fig. 18.5). To get an idea of the orders of magnitude, the temperature of the transition between the 3D and the 1D regimes for silicon is close to 1 K for a conductor of diameter $d = 100$ nm. For much smaller systems, such as carbon nanotubes, this universal quantum behaviour of K can be expected to arise at much higher temperatures. This value has been measured only once [20], and so far, few experiments have been able to measure the thermal conductance of a monolithic nanowire (good transmission coefficient with the reservoirs) at low temperatures [17].

Thermal Conductivity of Electrons

We shall be able to quote a large part of what has already been calculated for phonons, because in certain cases the carrier statistics has no effect and the results can be transposed without modification. The kinetic model can thus be applied to write

the electron thermal conductivity in the form

$$k_{e^-} = \frac{1}{3} C_{e^-} v_F \ell_{e^-} \ , \tag{18.25}$$

using (18.21). Then with the specific heat calculated in (18.10), we obtain the electron thermal conductivity as

$$k_{e^-} = \frac{\pi^2}{9} D(E_F) v_F \ell_{e^-} k_B^2 T \ . \tag{18.26}$$

In metals, the low temperature thermal conductance is largely dominated by this electronic contribution. Only superconductors have a low or even negligible electron thermal conductance due to the presence of a gap at the Fermi level.

Wiedemann–Franz Law

It is interesting to compare the expression for k_{e^-} in (18.26) with the expression for the electrical conductance [1], viz.,

$$\sigma = \frac{1}{3} D(E_F) v_F^2 \tau_{e^-} \ .$$

It turns out that the electrical and thermal conductivities are proportional. Charge transport clearly participates in heat transport:

$$\frac{k_{e^-}}{\sigma T} = \frac{\pi^2 k_B^2}{3e^2} = L_0 \ . \tag{18.27}$$

The coefficient L_0, which has the value $L_0 = 2.45 \times 10^{-8} \mathrm{W}\,\Omega\mathrm{K}^{-2}$, is called the Lorentz number. This relation is satisfied for many metals, over a wide temperature range. It is a very useful way of separating electron and phonon contributions when measuring the thermal conductivity of a metal.

Low-Dimensional Limit

The experimental demonstration of this limit is more delicate for electrons than for phonons. Indeed, the relevant characteristic lengths, such as the elastic mean free path ℓ_{e^-} (a few tens of nm) or the Fermi wavelength λ_F (less than nm) are generally much too small to give measurable finite size effects, at least in metals. However, in some systems such as 2D electron gases, e.g., the GaAs/AlGaAs heterostructure, the Fermi wave vectors can be much bigger. In this type of system, the quantization of *electrical* conductance has been observed [23] for quantum point contacts. The electrical conductance is then given by $G = 2e^2/h$ per conduction channel. Then by the Wiedemann–Franz law applied to a quantum point contact, the electron thermal

conductance is given by $k_{e^-} = 2\pi^2 k_B^2 T/3h$. We thus recover the universal value for the thermal conductance [see (18.24)]. The factor of 2 comes from the electron spin, whence two types of channel. It is thus remarkable that this universal value should be the same whatever the type of heat carrier and hence whatever the statistics of the heat carriers, be they boson or fermion [19].

18.3 Probing Thermal Properties by Electrical Measurements

In this section, we shall see how to obtain the value of a thermal quantity from experiments in which an electrical quantity is measured. We shall focus in particular on specific heat and thermal conductance measurements by steady state and dynamical methods. Their main advantage over other techniques is the possibility of making very sensitive measurements from a very precise measurement of a temperature change. The disadvantage is that direct contact must be made with the material under investigation, which may perturb the quantity being evaluated.

18.3.1 Thermometry

The key element in making a thermal measurement is of course the thermometer. For a microsystem, this will be a thin film. The quality of the temperature measurement depends on the temperature sensitivity of this element and the associated instrumentation. By definition, to be a good thermometer, a device must record some physical quantity that varies strongly with the temperature. In our case, this physical quantity will be an electrical quantity, e.g., a voltage, resistance, noise, etc. One of the limitations will arise precisely because we measure the temperature of the electron bath, which may in some cases be different from that of the phonon bath, depending on the interactions occurring in the two baths.

Thermometer Measurements

There are several basic techniques for obtaining a good temperature measurement T (K) by carrying out an electrical measurement. To begin with, one must always check that the current used for the measurement does not itself cause self-heating. Indeed, in the case of resistive thermometry, the measurement current dissipates heat by the Joule effect, and depending on the thermal coupling between the thermometer and the substrate whose temperature T (K) is to be measured, a thermal gradient may be set up between the two parts of the device. Such self-heating effects can considerably perturb the T (K) measurement, especially at low temperatures. Regarding the electrical connection with the thermometer, the most commonly applied technique uses the so-called four-wire setup (see Fig. 18.6a). This connection

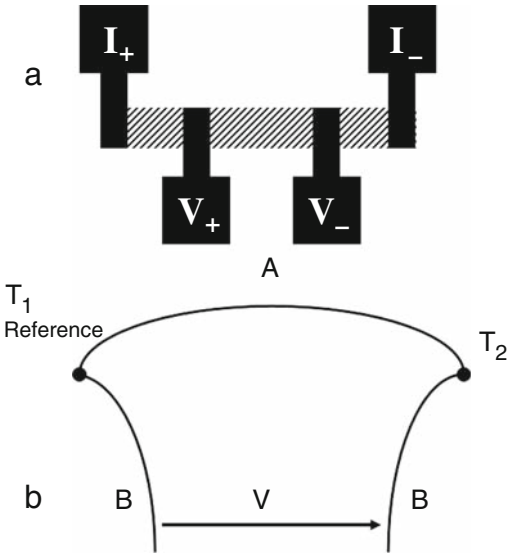

Fig. 18.6 (**a**) Four-wire transducer setup, with two external contacts (I_+ and I_-) for the measurement current and two internal contacts to measure the voltage, V_+ and V_-. (**b**) Thermoelectric effect. Two junctions are made between two materials A and B and held at different temperatures. A voltage V appears across the terminals of the two junctions

configuration allows spatial separation of the current and voltage leads, so that one extracts only the resistance of the thermometer, without including the contribution of the connection wires in the final value for the resistance. This configuration is very widely used in thermometry because it provides a highly reliable and reproducible, low noise measurement.

Thermocouple

One very widespread method of thermometry is based on thermoelectric effects, in particular, the Seebeck effect (see Fig. 18.6b). When two junctions (labelled 1 and 2) between two materials (A and B) are maintained at different temperatures T_1 and T_2, an electrical voltage $V = (S_B - S_A)(T_2 - T_1)$ is produced across the terminals of the circuit, where S_B and S_A are the Seebeck coefficients of the materials. Junction 1 is the temperature reference, and measurement of this voltage V will lead to a measurement of the temperature T_2.

The main advantage with this technique is that it is non-dissipative, because the thermometric element is not heated during the measurement, and there is no current circulating, hence no Joule effect. Unfortunately, at low temperatures, the Seebeck coefficients fall off rapidly, making it more difficult to use thermocouples.

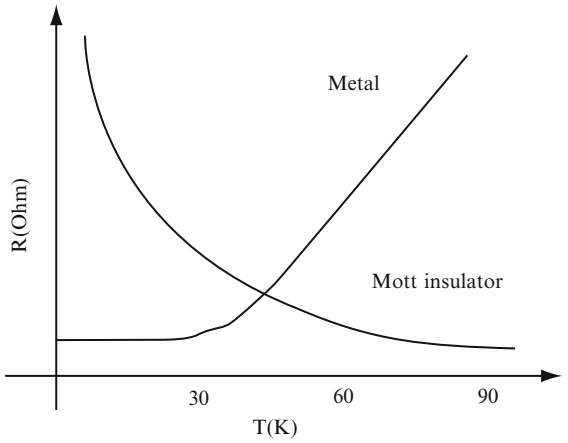

Fig. 18.7 Characteristics $R(T)$ for two types of resistive thermometer, one metal for which R is a linear function of T (positive temperature coefficient), and the second a Mott–Anderson insulator for which the resistance increases exponentially as T decreases (negative temperature coefficient)

Resistive Thermometry

Metals and semiconductors have a resistance that varies significantly with temperature. Resistive thermometry involves measuring the temperature by measuring the resistance of the thermometer, either calibrated previously with respect to a primary thermometer, or by a noise measurement, which works by definition as a primary thermometer. As shown in Fig. 18.7, metals have positive temperature coefficients $\alpha \approx 10^{-3}\ \mathrm{K}^{-1}$, where

$$\alpha = \frac{1}{R}\frac{\mathrm{d}R}{\mathrm{d}T}\ .$$

The resistance generally saturates around 30 K, so metals cannot be used at low temperatures. However, semiconductors and materials with a Mott–Anderson transition have negative temperature coefficients $\alpha \approx -1\ \mathrm{K}^{-1}$, making them particular effective at low temperatures. Indeed, these Mott–Anderson materials (NbSi [25], NbN [26]) become insulating when the temperature goes down far enough, whence the exponential increase in their resistance. For low temperature applications, thin films of Mott–Anderson insulator are thus among the most useful materials.

Other Low Temperature Thermometric Methods

Other types of thermometer exploit electrical measurements, particularly in low temperature applications, e.g., $S/I/N$ junction, Coulomb blockade junction. These types of thermometry has been described in detail by Giazotto et al. [27]. They lend themselves particularly well to integration in systems of very small dimensions, but cannot compete with resistance measurements when it comes to sensitivity.

Difficulties Specific to Low Temperatures

When a thermometer is biased with a measurement current I, care must always be taken over self-heating phenomena. Indeed, a power RI^2 is dissipated in the element and this is likely to perturb the temperature measurement. This phenomenon can become particularly problematic when working with small systems. The second problem that may arise comes from the fact that, at low temperatures, the electron–phonon relaxation time $\tau_{e^-\text{–ph}}$ becomes much longer than the electron–electron relaxation time. For this reason, when a temperature is to be measured by injecting electrical power, the temperature of the electrons will decouple from the temperature of the phonons in the same material. These relaxation times introduce a thermal gradient between the electron bath and the phonon bath that can be simply described by

$$T_{e^-} - T_{ph} = \frac{P_{e^-}}{K_{e^-/ph}} = \tau_{e^-\text{–ph}} \frac{V^2}{RC_{e^-}} \,. \tag{18.28}$$

This relation is only true for weak electric fields. Roukes et al. [28] have shown that, for strong fields, the relation between the two temperatures is given by

$$T_{e^-}^n - T_{ph}^n = \frac{P_{e^-}}{V g_{e^-/ph}} \,, \tag{18.29}$$

where n can take values up to 5, V is the volume, and $g_{e^-/ph}$ is the electron–phonon coupling constant, of the order of 1 000 $\mathrm{W/K^5cm^3}$ for metals. This means that, if a temperature is measured in these ranges ($T < 1$ K), these effects must be taken into account when the measurement parameters, e.g., V, I, frequency, etc., are chosen.

At the Nanoscale

When working with smaller and smaller objects, it is important to question the validity of the definition of temperature. Hartmann has discussed this problem for a 1D chain, and the orders of magnitude obtained can be somewhat surprising [29]. For silicon, the temperature cannot be defined for a sample measuring less than a few microns, which in 3D would give sizes of the order of a hundred or so nanometers at 1 K. However, this problem remains open. Moreover, reduced sample sizes have other thermodynamic consequences that need to be taken into account, e.g., the importance of fluctuations, energy competition between surface and volume, choice of the relevant statistical ensemble (conservation of particle number, microcanonical or canonical), and so on.

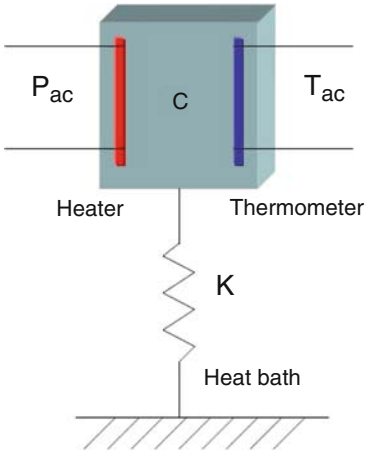

Fig. 18.8 Principle of calorimetric measurement. The measurement cell on which the transducers and sample are arranged is insulated from the thermal bath by a well defined thermal conductance K

18.3.2 Low Temperature Specific Heat Measurements at the Nanoscale

Up until the 1960s, specific heat measurements were generally made adiabatically. In this approach, the sample is considered to be completely isolated from the surroundings (thermal reservoir or heat bath). Any heat dissipation then causes a rise in temperature that will be proportional to the specific heat, i.e., $C = \delta Q/\delta T$. These methods are very efficient on bulk crystals, but prove difficult to apply to small systems, particularly when precise measurements are required for which it is difficult to obtain perfect thermal insulation. The measurement must be carried out in a short time τ_{meas} compared with the relaxation time toward the bath $\tau = C/K$ (see Fig. 18.8), and this is impossible to achieve for nanoscale samples. In the following, we shall merely describe the most commonly used techniques in low temperature nanoscience.

A.C. Calorimetry

Temperature modulated measurements, discovered at the beginning of the twentieth century by Corbino [30], have been resuscitated for high sensitivity applications in the work by Sullivan and Seidel [31], using what they call a.c. calorimetry. The reader is referred to the review by Kraftmakher for all temperature modulated calorimetry techniques [32].

The idea here is to heat with an alternating current in order to produce an oscillation in the temperature at a well defined frequency in the measurement cell (see Fig. 18.8). This oscillation will be detected by dynamical methods in the thermometer. Since it is difficult to obtain perfect thermal insulation for small systems, the

modulation is carried out at a higher frequency than the frequency corresponding to the characteristic time for heat loss to the bath, i.e., $1/\tau \ll f_{\text{meas}}$. However, there is a high frequency limit. Indeed, if the temperature oscillation is to be proportional to the specific heat, the whole system including thermometer, heating, and sample must be isothermal, i.e., the frequency of the temperature modulation must be less than the frequency corresponding to the characteristic diffusion time τ_{diff} in the system. One thus obtains an allowed frequency interval $1/\tau \ll f_{\text{meas}} \ll 1/\tau_{\text{diff}}$, within which the system can be considered to be quasi-adiabatic, i.e., effectively insulated from the thermal reservoir.

We shall now derive the various equations describing the temperature behaviour with this technique. We shall assume that the thermal contacts between the transducers, the membrane, and the sample are perfect, that this ensemble has a specific heat C, and is connected to the thermal bath by a thermal conductance K. A current $I = I_0 \cos(\omega t)$ is applied to the heater, dissipating a power $P_{\text{heat}} = RI_{\text{ac}}^2$. This power is at twice the frequency, so the temperature oscillation will also be at the frequency $2f$, as we shall see shortly. The heat balance is expressed in differential form, with the temperature rise being equated with the power supplied by the heater minus the power escaping into the thermal bath:

$$C\frac{\mathrm{d}T}{\mathrm{d}t} = P_{\text{heater}} - K(T - T_{\text{B}}) \,, \tag{18.30}$$

where

$$P_{\text{heater}} = \frac{RI_0^2}{2}\left[1 + \cos(2\omega t)\right]$$

is the power dissipated in the sensor by the Joule effect, and T_{B} is the temperature of the heat bath.

In order to remain within the linear case, the amplitude of the temperature oscillation δT_{ac} must be small enough to ensure that the various thermal parameters C, K, etc., can be treated as constant over this interval. With this assumption, the differential equation can be solved and one obtains the following solution:

$$T(t) = T_{\text{B}} + \frac{RI_0^2}{2K} + \delta T_{\text{ac}}^{\text{heater}} \,, \qquad \delta T_{\text{ac}}^{\text{heater}} = \frac{RI_0^2}{2\omega C}\left[1 + \frac{1}{(\omega\tau)^2}\right]^{-1/2} \cos(2\omega t + \varphi) \,. \tag{18.31}$$

The temperature $T(t)$ is thus a superposition of continuous terms, viz.,

$$T_{\text{DC}} = T_{\text{B}} + \frac{RI_0^2}{2K} \,,$$

and an alternating term $\delta T_{\text{ac}}^{\text{heater}}$ (see Fig. 18.9a). Taking into account diffusion in the setup between the two transducers (heater and thermometer), we obtain the following temperature in the thermometer (the temperature actually measured):

$$\delta T_{\text{ac}}^{\text{thermo}} = \frac{RI_0^2}{2\omega C}\left[1 + \frac{1}{(\omega\tau)^2} + (\omega\tau_{\text{diff}})^2\right]^{1/2} \cos(2\omega t + \varphi) \tag{18.32}$$

It is this component that carries information about the specific heat. Respecting the assumptions concerning the frequency, viz., $1/\tau \ll f_{\text{meas}} \ll 1/\tau_{\text{diff}}$, (18.32) becomes

$$\delta T_{\text{acRMS}}^{\text{thermo}} = \frac{RI_0^2}{2\omega C} \,. \tag{18.33}$$

So when a signal at frequency f is applied to the heater, we measure the heat capacity by measuring the voltage response at frequency $2f$ on the thermometer by standard lock-in amplifier methods.

Application to Nanocalorimetry on a Membrane

When we wish to measure very small samples, with masses less than a hundred nanograms, we implement a technique using a suspended and structured membrane (see Fig. 18.9b). Working with this kind of system, both the specific heat and the thermal conductance can be controlled so as to obtain the desired thermal behaviour (see Chap. 10). Several substrates are used at low temperatures, namely silicon and silicon nitride (Si_3N_4), as shown in Figs. 18.10 and 18.11. The first of these materials has the advantage of having a very low specific heat, while the second allows one to work with very thin membranes (thickness < 200 nm). Sensors with very small dimensions have been devised with silicon [33] and silicon nitride [34, 35], making the transducers by thin film deposition. The silicon devices were used to measure superconducting mesoscopic objects with submicron dimensions, for which specific phase transitions were revealed [36, 37]. The energy resolution of these sensors (see

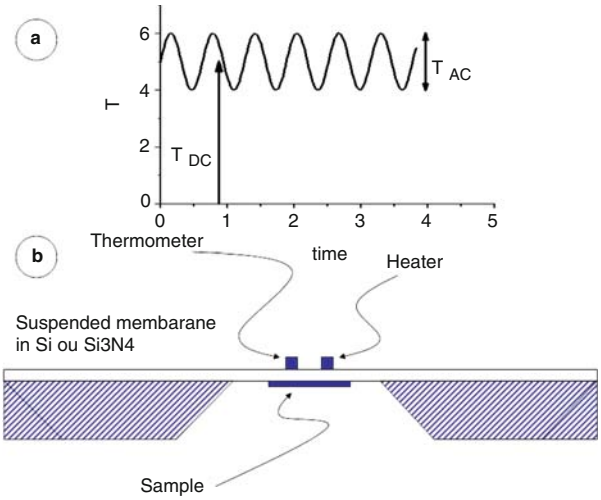

Fig. 18.9 (**a**) Temperature profile of a membrane. A continuous temperature gradient T_{DC} is superposed on a temperature oscillation δT_{ac} with amplitude inversely proportional to the heat capacity C of the cell. (**b**) Suspended membrane nanocalorimeter

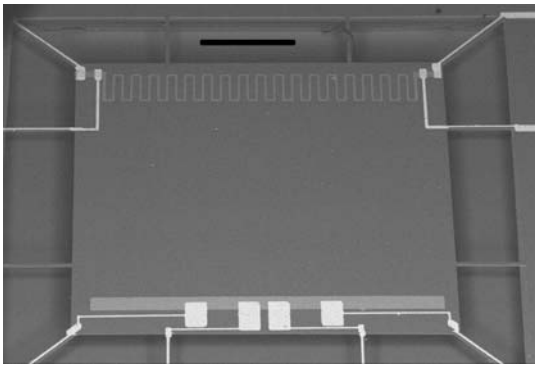

Fig. 18.10 Scanning electron microscope (SEM) image of a silicon membrane sensor. Two transducers are visible, the copper heater at the top and the NbN thermometer at the bottom. *Scale bar* 1 mm [36, 38, 39]

Fig. 18.10) is a few attojoule (10^{-18} joule). The sensitivity per μm^2 is a few k_B, which opens fascinating prospects for measuring objects with dimensions of a few nanometers [38, 39]. Finally, note that other techniques have been developed for measuring microcrystals with a.c. heating by light-emitting diode [40].

Relaxation Calorimetry

This technique can be implemented by the same kind of setup as the one shown in Fig. 18.8. The temperature of the membrane is held at $T = T_0 + \delta T$. Then, the heating of the membrane is cut off in a very short time and the temperature relaxes to that of the bath via an exponential decay of the form

$$T(t) = T_0 + \delta T e^{-t/\tau} , \tag{18.34}$$

where $\tau = C/K$, as for a.c. calorimetry. This relaxation can occur over a very short time scale, of the order of a hundred or so microseconds. This is therefore a faster measurement technique than a.c. calorimetry. Having measured K by considering a steady state, i.e., δT constant, the specific heat can then be deduced. This technique was the one used to measure C_P on an Si_3N_4 membrane by Hellman and coworkers [34] and by Roukes et al. [35], who were the first to make a sensor on the μm scale (see, for example, Fig. 18.11). Unfortunately, these devices are not generally sensitive enough to envisage measurements on nanoscale objects, either because it is not a temperature modulated technique, or because the thermometry used does not involve high enough temperature coefficients. However, for measurements on thin films over broad temperature ranges, e.g., 2–300 K, this technique has few equals [34].

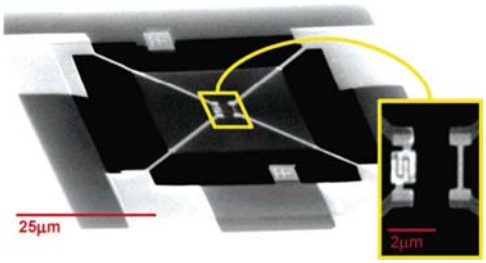

Fig. 18.11 SEM image of a silicon nitride membrane sensor. The two transducers are visible at the *center*, made by electron lithography: AuGe for the thermometer (*left*) and Au (*right*) for the heater. The membrane measures 25 μm across [35]

18.3.3 Thermal Conductance Measurements on Nanoscale Samples

Very few experiments provide a way of measuring the thermal conductance on systems with nanoscale dimensions. Of course, the difficulty lies in the thermal connection of these samples with the bath, but also in the measurement itself. The coupling between the nanosystem and the thermometric element must be made very carefully in order to control temperature variations induced by a heater. The basic thermal conductance measurement involves supplying heat power to one end of the system and measuring the resulting temperature gradient: $K = P/\Delta T$. However, the various methods can be divided into two groups:

- Steady-state methods, in which the temperature gradient is continuous.
- Dynamical methods, in which the thermal gradient varies with time and which allow a frequency detection of the electrical signal.

Steady-State Methods

These methods can be implemented in a very direct way. However, we shall see that interpretation of the results can prove difficult. So for some applications, dynamical methods are preferable, especially, thin films and thick layers. One approach is the Völklein method [41], where the thermal conductance is measured on a membrane. To do this, a thermometer with small width is deposited lengthwise on a membrane with very high aspect ratio. By the Joule effect, this thermometer is heated by its own measurement current, and the resulting temperature rise is recorded in order to deduce the value of the lateral thermal conductance (see Fig. 18.12a) by solving the heat transfer equations between the center (the hottest point) and the silicon frame (thermal reference).

For measurements at much lower temperatures, one of the chosen configurations is shown in Fig. 18.12b. The idea is to measure the thermal conductance of the suspensions of a self-supporting membrane. Using micro- and nanofabrication

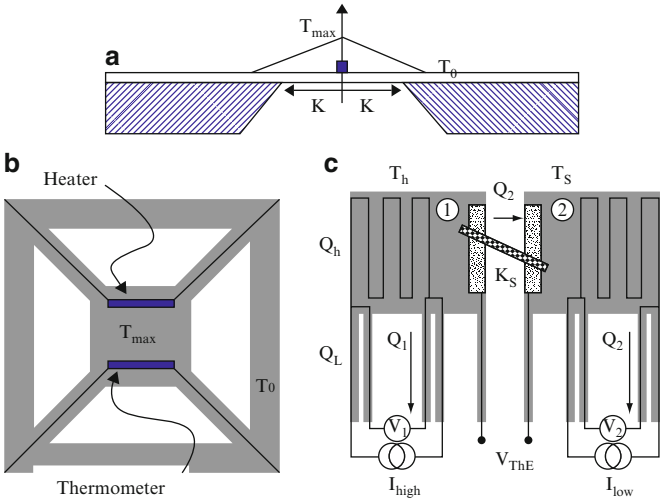

Fig. 18.12 Different types of thermal conductance measurement. (**a**) Völklein arrangement. A thermal gradient is set up between the center of the membrane and the frame. (**b**) This arrangement can be used to measure the thermal conductance of the suspensions. (**c**) Arrangement due to L. Shi et al. [42]. A thermal gradient is set up between membranes 1 and 2, with a platinum thermometer structured on each. The sample S connects the two membranes

methods, a thermometer and heater are deposited at the center. By measuring the thermal gradient, one can then deduce the thermal conductance of the nanostructured support. This method can be used to carry out measurements on structures with widths a few tens of nanometers and lengths in the micron range (nanowires). With this technique, M. Roukes and coworkers were able to reveal 1D thermal transport effects in silicon nitride membranes [20]. This is one of the rare results confirming universal heat transport behaviour, as discussed in Sect. 18.2.2 (see p. 548). The advantage with this technique is to be able to work on monolithic systems, and hence to eliminate the thermal contact resistances between the nanowire and its reservoirs. However, the main drawback lies in the presence of interfering thermal conductors on the surface of the wires, in this case the current leads. For temperatures above $T_c/10$, even if the current leads are superconducting, they still conduct heat by non-condensed electrons at a significant level compared with a dielectric.

Another method has been devised by L. Shi and coworkers [42], illustrated in Fig. 18.12c. The setup comprises two membranes, each carrying a platinum thermometer. The sample, e.g., a nanowire, is suspended between the two membranes. This system can be used to find the thermal gradient between the two membranes without having electrical connections at the surface of the object under investigation. Each thermometer is read using the four-wire arrangement. Data interpretation is complex because the power Q_2 actually crossing the sample must be precisely determined. Indeed, a part Q_1 of the total power is evacuated by the membrane suspensions. This power Q_1 can be evaluated by taking a measurement without the sample. The conductance K_S of the sample is then found from the relation

$$K_S = K_{susp} \frac{\Delta T_S}{\Delta T_h - \Delta T_S} , \tag{18.35}$$

where K_{susp} is the thermal conductance of the membrane suspensions toward the heat bath, and ΔT_h, ΔT_S are the thermal gradients between the membranes and the heat bath.

In addition, and this is one of the main advantages of this technique, by virtue of two platinum electrodes on the sample (the two central electrodes V_{ThE} in Fig. 18.12c), this arrangement can be used to make a direct measurement of the Seebeck coefficient S of the nanowire. Indeed, the thermoelectric properties of samples structured on a very small scale is one of the main themes of heat transfer physics today. The experimental and theoretical challenge is to reduce phonon thermal conductance while maintaining a high electrical conductance [43, 44].

This technique has been used to carry out many thermal conductance measurements on nanowires (above 10 K), and also on carbon nanotubes or silicon nanowires (see the review [45]). The important result is the elimination, in wires of very small cross-section (diameter less than 30 nm), of the thermal conductivity peak due to Umklapp processes. This behaviour, very different from what happens in the bulk, would be the signature of the dominant presence of scattering at the surfaces of the nanowires [43,46]. Phonon transport in low dimensions would be dominated by what happens at the surfaces: competition between specular reflection and scattering. However, the weak point with this technique comes from contact thermal resistances between the nanowire and membranes, these being poorly controlled.

Dynamical Methods

In the last section, the methods described were based on the idea of setting up a steady state thermal gradient. The sensitivity of a thermal conductance measurement can be greatly improved by using a method in which the thermal gradient is time dependent. These methods can be compared with a.c. calorimetry techniques, where the temperature is modulated in time.

The 3ω Method. Radial Measurements

The idea here is to heat a wire, which serves both as thermometer and heater, by an a.c. current in order to produce a radial heat flux in the material to be studied. The experimental arrangement is shown in cross-section in Fig. 18.13a. As in a.c. calorimetry, a current $I = I_0 \cos(\omega t)$ is passed through the transducer. The temperature oscillates in the same way at $2f$ (see p. 555). The amplitude of the oscillation will depend directly on the thermal conductance of the material under the transducer. The voltage signal

$$V = \left\{ R_0 + \frac{dR}{dT} \left[T(t) - T_0 \right] \right\} I_0 \cos(\omega t)$$

across the thermometer terminals includes a term at f, related to $V = R_0 I_0$, and a term at $3f$ arising from

$$V = \frac{\mathrm{d}R}{\mathrm{d}T}\Delta T \cos(2\omega t + \varphi) I_0 \cos(\omega t) \; .$$

To find the term in ΔT, one must solve the heat transfer equations corresponding to the sample geometry. For the case depicted in Fig. 18.13a, a temperature oscillation arises in the positive half-plane (x, y), where the heat diffusion wave has wavelength $\lambda_\mathrm{h} = \sqrt{K/\mathrm{i}2\omega\rho C}$. The solution to this equation in cylindrical coordinates is [47]

$$\Delta T(r) = \frac{R I_0^2}{l\pi K} J_0\big(r/\lambda_\mathrm{h}\big) \; , \tag{18.36}$$

where K is the thermal conductivity of the half-volume under the transducer and J_0 is the zero order Bessel function [48]. This method can thus be used to measure the thermal conductivity of a thick film, treated as infinite as compared with the thermal wavelength λ_h and attenuation. As can be seen from (18.36), this method also measures the specific heat, and this was indeed its first application [49].

The 3ω Method. Experimental Setup

The $3f$ component of the transducer voltage is recorded by a measurement chain using a lock-in amplifier technique. The main drawback of the 3ω method is the presence of a significant f component proportional to $R_0 I_0$, which is impossible to filter using passive systems. It is thus advisable to mount the transducer in a Wheatstone bridge configuration in order to eliminate the f component [49] and hence carry out the measurement in the best instrumental conditions. The sensitivity of the measurement is limited by the thermal noise of the transducer and is directly proportional to the temperature coefficient

$$\alpha = \frac{1}{R}\frac{\mathrm{d}R}{\mathrm{d}T}$$

of the thermometer. Compared with other ways of measuring the thermal conductivity, this has the best signal-to-noise ratio.

The 3ω Method. Longitudinal Measurement on Nanowires

One recent application of the 3ω method carries out direct measurements on mechanically suspended nanowires (carbon nanotubes and silicon nanowires). If we consider a suspended 1D system with a thermally insulating transducer on the surface, we can produce a temperature oscillation between the center of the wire and the edges, which serve as heat bath, by applying an a.c. current across the transducer terminals. This arrangement suggested by Lu et al. is depicted in Fig. 18.13b [50]. In contrast

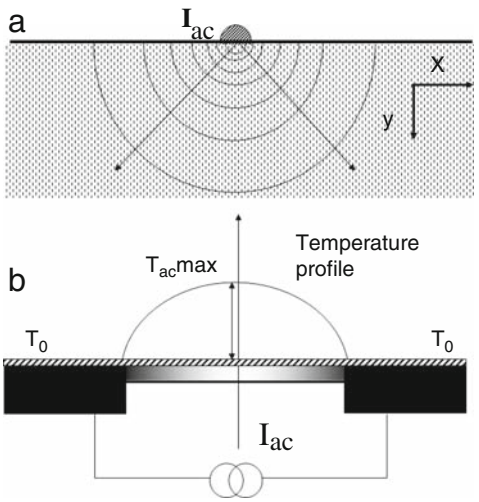

Fig. 18.13 (**a**) The 3ω method for a cross-plane thermal conductivity measurement. The thermometer is shown in cross-section and *concentric rings* represent heat waves moving out normally to the surface. (**b**) The 3ω method for a thermal conductivity measurement along a nanowire. *Shaded* material represents the thermometer deposited on the black substrate

to the arrangement chosen by Cahill, where the thermal conductivity is measured radially, this setup measures the longitudinal thermal conductivity. In the configuration of Fig. 18.13b, as for the other dynamical methods, a current $I = I_0 \cos(\omega t)$ is applied across the transducer terminals, and the heat equations must therefore be solved for a 1D system. The power balance in the system is given by

$$C\frac{\partial T(x,t)}{\partial t} = I_0^2 \cos^2(\omega t)\left\{R + \frac{\mathrm{d}R}{\mathrm{d}T}\left[T(x,t) - T_0\right]\right\} - K\frac{\partial^2 T(x,t)}{\partial x^2}\,, \qquad (18.37)$$

where C is the heat capacity of the nanowire and K its thermal conductance. We prefer to use these physical data depending on the geometry of the material because, as we have seen, at these length scales, the very concept of thermal conductivity does not necessarily have any meaning. The last equation simply expresses the fact that any increase in temperature over a time t corresponds to the power dissipated in the transducer minus the power evacuated by heat loss to the bath at temperature T_0. To solve this differential equation, we shall take as boundary conditions that the temperature is given by $T = T_0$ at the points of contact $x = \pm L/2$ with the heat bath, T_0 being fixed by the reservoirs. By symmetry, given that the Joule power is dissipated right along the wire, the thermal gradient arises between the center of the nanowire and the reservoirs.

To solve the differential equation, we need to make some assumptions about the system. To remain within the linear approximation, the total dissipated power must be much greater than the non-uniformities in the power dissipated due to temperature variations along the wire. This is expressed by the relation $\Delta P \ll P_0$, which

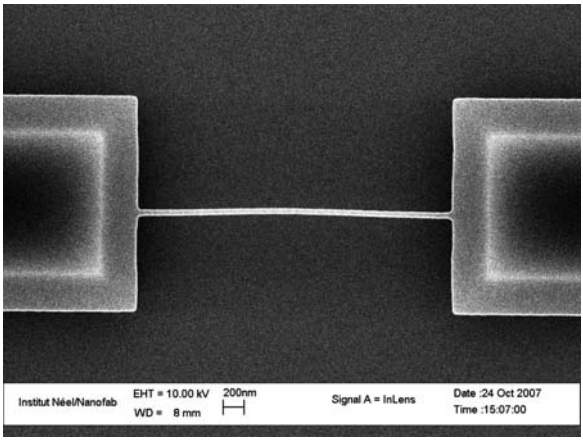

Fig. 18.14 SEM image of a silicon nanowire of width 50 nm, connected to two reservoirs

implies

$$\frac{\mathrm{d}R}{\mathrm{d}T} I_0^2 \Delta T \ll R I_0^2 \,,$$

and hence, $\alpha \Delta T \ll 1$, where α is the temperature coefficient of the transducer [50]. This assumption is well confirmed, even at low temperatures, in the experimental situations we are discussing (see Fig. 18.14) [17].

The solutions to this equation are, to first order,

$$V_{3\omega} \approx \frac{2 I_0^3 R^2 \alpha}{\pi^4 K \sqrt{1 + (2\omega\gamma)^2}} \sin(3\omega t - \varphi) \,, \tag{18.38}$$

where $\tan\varphi \approx 2\omega\gamma$ gives the phase difference between the current excitation and the voltage response. The limit that interests us here is the low frequency limit, occurring when $f < 1/\gamma$. The value of the voltage at 3ω rms is given by

$$V_{3\omega} \approx \frac{4 I_0^3 R^2 \alpha}{\pi^4 K} \,. \tag{18.39}$$

In this case, the voltage measurement at the third harmonic directly delivers the thermal conductance K of the nanowire plus transducer system. The method assumes that the temperature can be defined at each point of the wire, something which eventually turns out to be false at very low dimensions.

Thermal conductance measurements on monolithic silicon wires have been carried out using this method. Reduced size effects have been demonstrated, as can be seen in Fig. 18.15, where, below 1.2 K, the thermal conductance deviates from the T^3 law, the signature of edge effects on K.

Note that in the very high frequency limit, this experimental technique has also been applied to measure the specific heat of suspended nanowires [50].

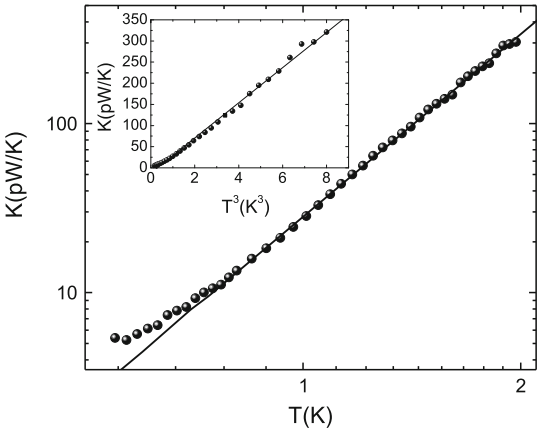

Fig. 18.15 Thermal conductance results for a silicon nanowire at low temperature. *Insert*: The thermal conductance is plotted as a function of the third power of the temperature to illustrate the T^3 law [17]

18.4 Conclusions

We have described as fully as possible the various techniques available for low temperature heat transfer physics. Depending on the application and depending on the geometry of the system, there is a wide choice of methods. Some areas have been deliberately omitted, such as bolometric methods which have very different scientific objectives and which are exhaustively described in other review articles [27]. Despite the intrinsic difficulties at low temperatures, the main advantages in probing thermal properties by electrical methods stem from the very precise control over injected power and a very accurate measurement of temperature changes through high sensitivity thermometry.

A great many open questions remain concerning thermal problems on small scales, particularly at low temperatures. There are few experimental results and experiments are generally difficult to implement. As examples, we may cite:

- Demonstration of the existence of a gap in the phonon density of states for structured nanowires by observation of a minimum in the thermal conductance [51].
- Implementation of time-resolved heat transfer experiments for single-phonon detection [52].
- Measurement of the heat capacity of a single nanoscale object.
- Development of a new generation of thermometry able to improve coupling between electron and phonon baths in nanosystems.
- Improvement in the efficiency of non-dissipative forms of thermometry.

Fascinating prospects are thus opening up in this area of thermal nanophysics. The convergence with other types of heat transfer technique may lead to important progress with regard to the questions raised today, despite the specific experimental constraints imposed by low temperature physics.

Acknowledgements I would particularly like to thank my colleagues J. Chaussy, T. Fournier, J.-L. Garden, H. Guillou, J.-S. Heron, F. Ong, J. Richard, G. Souche, and all the technical staff of Nanofab and Pôle Capteur Thermométrique et Calorimétrie.

References

1. N.W. Ashcroft and N.D. Mermin: *Solid State Physics* (Saunders College Publishing, New York, 1976)
2. H.M. Rosenberg: *Low Temperature Solid State Physics* (Clarendon Press, Oxford, 1963)
3. F. Pobell: *Matter and Methods at Low Temperatures* (Springer-Verlag, Berlin, 1992)
4. J.M. Ziman: *Electrons and Phonons* (Clarendon Press, Oxford, 2001)
5. E.S.R. Gopal: *Specific Heats at Low Temperatures* (Clarendon Press, Oxford, 2001)
6. A. Tari: *Specific Heat of Matter at Low Temperatures* (Imperial College Press, London, 2003)
7. G. Chen: *Nanoscale Energy Transport and Conversion* (University Press, Oxford, 2005)
8. B. Diu, C. Guthmann, D. Lederer, B. Roulet: *Physique Statistique* (Hermann Editeurs, Paris, 2001)
9. D.V. Anghel, J.P. Pekola, M.M. Leivo, J.K. Suoknuuti, and M. Manninen: Phys. Rev. Lett. **81**, 2958 (1998)
10. J. Hone, B. Batlogg, Z. Benes, A.T. Johnson, and J.E. Fischer: Science **289**, 1730 (2000)
11. A. Gusso and L.G.C. Rego: Phys. Rev. B **75**, 045320 (2007)
12. W. Yi, L. Lu, Z. Dian Lin, Z.P. Pan, and S.S. Xie: Phys. Rev. B **59**, R9015 (1999)
13. L.X. Benedict, S.G. Louie, and M.L. Cohen: Solid State Commun. **100**, 177 (1996)
14. K.S. Yi, D. Kim, and K.-S. Parks: Phys. Rev. B **76**, 115410 (2007)
15. W. Schottky: Phys. Z **23**, 448 (1922)
16. D.H. Santamore and M.C. Cross: Phys. Rev. Lett. **87**, 115502 (2001)
17. O. Bourgeois, T. Fournier, and J. Chaussy: J. Appl. Phys. **101**, 016104 (2007)
18. E.T. Swartz and R.O. Pohl: Rev. Pod. Phys. **61**, 605 (1989)
19. L.G.C. Rego and G. Kirczenow: Phys. Rev. Lett. **81**, 232 (1998)
20. K. Schwab, E.A. Henriksen, J.M. Worlock, M.L. Roukes: Nature **404**, 974 (2000)
21. J.B. Pendry: J. Phys. **16**, 2161 (1983)
22. R. Maynard and E. Akkermans: Phys. Rev. B **32**, 5440 (1985)
23. B.J. van Wees, H. van Houten, C.W.J. Beenakker, J.G. Williamson, L.P. Kouwenhoven, D. van der Marel, C.T. Foxon: Phys. Rev. Lett. **60**, 848 (1988)
24. N. Nishiguchi, Y. Ando, M.N. Wybourne: J. Phys. Condens. Matter **9**, 5751 (1997)
25. D.W. Denlinger, E.N. Abarra, K. Allen, P.W. Rooney, M.T. Messer, S.K. Watson, F. Hellman: Rev. Sci. Instrum. **65**, 946 (1994)
26. O. Bourgeois, E. André, C. Macovei, and J. Chaussy: Rev. Sci. Instrum. **77**, 126108 (2006)
27. F. Giazotto, T.T. Heikkila, A. Luukanen, A.M. Savin, J.P. Pekola: Rev. Mod. Phys. **78**, 217 (2006)
28. M.L. Roukes, M.R. Freeman, R.S. Germain, R.C. Richardson, M.B. Ketchen: Phys. Rev. Lett. **55**, 422 (1985)
29. M. Hartmann, G. Mahler, O. Hess: Phys. Rev. Lett. **93**, 080402 (2004)
30. O.M. Corbino: Phys. Z. **11**, 413 (1910); ibid. **12**, 292 (1911)
31. P.E. Sullivan, G. Seidel: Phys. Rev. **173**, 679 (1968)
32. Y. Kraftmakher: Phys. Rep. **356**, 1 (2002)
33. F. Fominaya, T. Fournier, P. Gandit, and J. Chaussy: Rev. Sci. Instrum. **68**, 4191 (1997)
34. B. Revaz, B.L. Zink, F. Hellman: Thermochimica Acta **432**, 158 (2005)
35. W. Chung Fon, K.C. Schwab, J.M. Worlock, M.L. Roukes: Nano Lett. **5**, 1968 (2005)
36. O. Bourgeois, S.E. Skipetrov, F. Ong, J. Chaussy: Phys. Rev. Lett. **94**, 057007 (2005)
37. F.R. Ong, O. Bourgeois, S.E. Skipetrov, J. Chaussy, S. Popa, J. Mars, and J.-L. Lacoume: Physica C **466**, 37 (2007)

38. F.R. Ong, O. Bourgeois, S.E. Skipetrov, and J. Chaussy: Phys. Rev. B **71**, 140503(R) (2006)
39. F.R. Ong, O. Bourgeois: Europhys. Lett. **79**, 67003 (2007)
40. M.B. Salamon, S.E. Inderhees, J.P. Rice, B.G. Pazol, D.M. Ginsberg, and N. Goldenfeld: Phys. Rev. B **38**, 885 (1988)
41. F. Völklein: Thin Solid Films **188**, 27 (1990)
42. L. Shi, D. Li, C. Yu, W. Jang, D. Kim, Z. Yao, P. Kim, A. Majumdar: J. Heat Transfer **125**, 881 (2003)
43. A.I. Hochbaum, R. Chen, R.D. Delgado, W. Liang, E.C. Garnett, M. Najarian, A. Majumdar, and P.Yang: Nature **451**, 163 (2008)
44. A.I. Boukai, Y. Bunimovich, J. Tahir-Kheli, J.-K. Yu, W.A. Goddard III, and J.R. Heath: Nature **451**, 168 (2008)
45. D.G. Cahill, W.K. Ford, K.E. Goodson, G.D. Mahan, A. Majumdar, H.J. Maris, R. Merlin, and S.R. Phillpot: J. Appl. Phys. **93**, 793 (2003)
46. N. Mingo, L. Yang, D. Li, and A. Majumdar: Nano Lett. **3**, 1713 (2003)
47. D.G. Cahill: Rev. Sci. Instrum. **61**, 802 (1990)
48. H.S. Carslaw and J.C. Jaeger: *Conduction of Heat in Solids* (Clarendon Press, Oxford, 1967)
49. N.O. Birge and S.R. Nagel: Rev. Sci. Instrum. **58**, 1464 (1987)
50. L. Lu, W. Yi, and D.L. Zhang: Rev. Sci. Instrum. **72**, 2996 (2001)
51. A.N. Cleland, D.R. Schmidt, C.S. Yung: Phys. Rev. B **64**, 172301 (2001)
52. M.L. Roukes: Physica B **263**, 1 (1999)

Index

ab initio method, 92, 392
absorbance, 470, 472
  red-edge, 331
absorption band, 439, 470, 472, 473, 476, 487
absorption coefficient, 128, 151
  of nanoparticle ensemble, 162
absorption cross-section, 506
  of organic molecules, 508
  of oscillating dipole ensemble, 184
  temperature dependence, 507
  two-photon, 507
absorption efficiency, 160, 169
absorption line, 439
  ytterbium, 509
acoustic mismatch model, 547
acridine orange, 323
acridine yellow, 312
active layer, 373, 374, 403
active targeting, 346–347
  magnetic, 347
aerogel, 5
AIDS, 353
Airy disk, 441, 442
Airy pattern, 441, 442
Allen–Feldman method, 63
alopecia, 347
Alumel wire, 429, 430
alumina matrix, 141, 142
aluminium, 256, 371, 392
  emissivity, 416
  resistance, 406
  thermistance, 401
  track, 399, 404
aluminium nitride, 377, 384
aluminium oxide strip, 522–524
  fluorescence image, 523
amide I, 499
aminosilane, 360
amphotericin, 350
amplifier, 379
  broad band, 380
analytical sciences, 302
  miniaturisation, 302
angiography, 349
anharmonic interaction, 77
anthracene-anisidine, 312
anthrax, 302
anti-Stokes excitation, 331, 334, 509
antibody, 347
apoptosis, 353
arterial embolisation, 359
association rate constant, 305
atmosphere, 152, 168
atom trapping, 286
atomic force microscope, 440, 445, 469
  cantilever, 444, 476, 482–486
  contact mode, 482–485
  for SNOM, 449
  for SThM, 518
  piezoresistive cantilever, 533
  tapping mode, 452
  thermal, 411
  to detect thermal expansion, 482–485
  to measure absorption, 485–488
  topographic image, 444
  tungsten tip, 517
autocorrelation function, 45
automation, 301

backscatter interferometry, 310
bacteria, 473, 492, 493, 499
  FTIR absorption spectrum, 497
  PTIR absorption spectrum, 497
bacteriophage, 493, 495, 496, 500

ballistic heat transfer, 9, 10, 24, 29–30, 52
  in nanofilm, 50
ballistic–diffusive equation, 142–146
band bending, 234, 241
Bardeen approximation, 76
BBD, 424
BCECF, 328
Beer–Lambert law, 470
Bessel functions, 164, 206, 209, 562
binomial distribution, 419
bioavailability, 346
biocompatibility, 349, 352
biodistribution, 344, 351
  monitoring, 352
biological analysis, 301, 334
  by SNOM, 447, 454
biological labelling, 343
biological microsystem, 370
biomedical imaging, 131
birefringence, 315
black body, 415–417, 457, 459
  double cavity, 429, 430
  phonon, 32, 38, 541, 547
  thermal emission, 416
  total emission, 417
Bloch wall, 355
blood, 302
blood clearance, 344, 345
blood compartment, 343, 344, 346
blood–brain barrier, 346, 349
Blue Tooth, 380
Boltzmann distribution, 330, 334, 507, 510, 511
Boltzmann transport equation, 10, 15, 20, 22–24, 31, 50, 92, 115, 130, 137, 141, 142, 144, 234, 237, 238, 247, 281, 545
  collision integral, 545
  convective term, 23
  electron–electron scattering term, 138
  electron–phonon scattering term, 138
  in cylindrical geometry, 32
  linear approximation, 27
  linearised, 30
  molecular dynamics, 40–49
  radiative formulation, 23–24
  scattering term, 23, 138, 545
  solution, 139
  source term, 138
  spatial discretisation, 34
  spectral discretisation, 35
  temporal discretisation, 35
  with Monte Carlo method, 33–40
bone marrow, 344, 345
Born–Von Karman model, 11, 20
boron nitride, 376, 377, 379, 384
Bose–Einstein distribution, 21, 23, 70, 115, 138, 143, 459, 464, 539
Bragg reflection, 314
brain, 346
breakdown voltage, 378
Brenner interaction potential, 44
Brillouin zone, 13, 129
  first, 18, 26, 35, 539
  L point, 130
Brownian conditions, 9, 214
Brownian relaxation, 355–357
buffer, 302, 328, 329, 332, 335
bulk medium hypothesis, 21
buried object, 390, 391, 478, 480, 489, 491, 500–502

caloduct, 370, 384
calorimetry, 305
  a.c., 555–558
  isothermal, 306
  miniaturisation, 306
  nanoscale, 557, 558
  on membrane, 557
  relaxation, 558
  sensitivity, 558
  temperature modulated, 555–558
cancer, 347, 351, 362
  thermal treatment, 17, 127, 350, 353
Candida albicans, 492
  blastospore, 493
  chemical mapping, 498
  glycogen band, 498
  hypha, 493, 495, 498
  IR absorption spectra, 496
  optical image, 493
  PTIR study, 495
capillary tube, 302, 303, 335
  temperature profile in, 303
cardiovascular system, 343
Carnot limit, 244, 247, 251, 252, 283
Casimir theory, 546
Cattaneo–Vernotte model, 144
CCD, 424–425
  dynamic range, 424
  physical principle, 424
  structure, 425
  transfer inefficiency, 425
CCD camera, 390, 391, 393, 395, 415, 425
  dynamic range, 431
  silicon, 430–435
  thermography, 431
CCD spectrometer, 395
ceramic, 330

doped, 332
chemical mapping, 498–501
of Candida albicans, 498
of E. coli, 499, 501
of T5 virus, 500
chemical vapour deposition, 3, 255
chemotherapy, 347, 350, 353, 361
cholesteric liquid crystal, 309, 314–315
chromophore, 328
Clausius–Mossotti model, 187–188, 192
clinical trials, 344, 359–360
cloud, 152
radiative response, 173–175
CMOS, 367, 373, 376, 381, 424
coarse-graining, 92
coil–globule transition, 315
colloid, 311, 316
complexation, 325
composite medium, 95
composite substrate, 377
COMSOL, 477
conductance quantum, 30
confinement, 10, 11, 24, 28, 50–56, 131, 237, 241, 243, 253, 254, 538
of EM field, 442–445
confocal microscopy, 473–475, 498
contrast agent, 349, 360
ligand-bearing, 349
magnetic susceptibility, 349
paramagnetic, 349
plasma half-life, 349
convective heat transfer, 17
cooling, 6, 213, 218, 226–229, 236, 244, 251, 256, 258, 261, 277, 281, 372, 378, 379
by Peltier effect, 226, 227, 235, 274, 281, 285, 426
by superlattice, 256, 258, 260–263
by thin film, 255–257
coefficient of performance, 228
cryogenic, 247
fluid-based, 228, 272
HIT, 246, 247, 261, 284
in telecommunications, 380
localised, 256, 274
nanoscale, 243
of chip, 261, 274
of integrated circuits, 274, 390
of laser diodes, 228, 384
of transistors, 377
power density, 245, 246, 259–261, 273, 274
solid-state, 228, 230
thermionic emission, 246, 286
WDM system, 273
coordination sphere, 329
copper, 128, 371
interface with liquid helium, 548
interface with silicon, 548
nanoparticle, 213
surface plasmon resonance, 131
core–shell particle, 145–147
silica/gold, 351
Coulomb blockade junction, 553
crystal field, 506, 507
crystal lattice, 18, 538, 542
defects, 24
eigenmodes, 21
silicon, 19
sub-wavelength spatial variations, 454
cyclodextrin complexation, 327–328

data processing, 370, 374, 376, 381
data storage, 368, 370, 374, 376, 381
DBD, 316
Debye angular velocity, 116
Debye approximation, 14, 22, 34, 116, 139, 540–541
Debye speed, 116, 125
Debye temperature, 20, 21, 117, 125, 514, 541
high, 542
Debye wave vector, 540
decimation technique, 67
density, 378
density of states, 116, 242
electromagnetic, *see* EM-LDOS
electron, 543
phonon, 12, 22, 143, 546
spectral, 65, 72, 76, 86
dependent radiation–matter interaction regime, 171, 172
dextran, 346, 349, 360
corona, 349
diamond, 374–376, 379, 384
doped, 375
enriched, 372
in ZnS, 113, 114
insulating layer, 377
monocrystalline, 374
polycrystalline, 374
thermal conductivity, 3, 374
wafer, 374, 375
dielectric contrast, 449–451
mapping, 454
dielectric function, 128
of noble metals, 128–130, 133
differential conductivity, 232
diffraction, 167, 439–442
diffuse mismatch model, 30, 547
diffusion-limited aggregation, 201, 202

diffusion-limited cluster–cluster aggregation, 201, 202
diffusive heat transfer, 10, 17, 38, 52
  in dielectric matrix, 141
  in nanofilm, 50
digital alloy, 266
digital photography, 370
discrete dipole approximation, 175–204
  absorption cross-section, 184
  applications, 191–204
  assumptions, 177
  calculating induced dipole moments, 178
  comparison with Mie theory, 191–197
  DDSCAT, 176
  extinction cross-section of oscillating dipole ensemble, 181
  polarisability models, 187–191
  scattering by oscillating dipole ensemble, 180
  scattering cross-section of oscillating dipole ensemble, 181
  scattering phase function, 183
  theory, 176–187
discrete ordinate method, 27, 31, 50, 195
  cylindrical geometry, 31–33
disordered alloy, 120–121
dispersion, 12, 28, 35, 79, 130
dispersion relation, 19–21, 29, 42, 540
  graphite, 79
  linear, 540, 542
  silicon, 20
  SWNT, 77
dissociation rate constant, 305
distributed feedback laser, 273
DNA
  amplification by PCR, 303–305
  double helix, 304
  factory, 493
  molecular beacon, 326
  phosphate band, 496, 499, 500
  sequencing, 301, 303
  viral, 494
DNA–RhG, 312
donor–acceptor pH probe, 329
doping, 25, 330, 332, 372–374, 390, 454, 508
  boron, 375
  effect on thermal conductivity, 120, 372
  erbium, 515, 527
  europium, 511
  phosphor, 375
  rare earth ion, 509–511
  transition ion, 512–513
  ytterbium, 515
Doyle model, 190
DPP, 312, 322
Draine model, 189–190, 192, 195
Drude model, 128, 136, 392, 418
drug delivery, 17, 343, 362
  active targeting, 346–347
  passive targeting, 345
  with controlled release, 350, 352
drug discovery, 305
dynode, 423
Dyson equation, 66, 68, 87, 89

eddy current, 354, 357
effective medium model, 110, 113, 114, 122–124, 132, 133, 136, 214
  for nanoparticle suspension, 217
  for nanotube suspension, 217
  limitations, 114–115
  multilevel, 218
electrical conductance, 550
electrical conductivity, 227, 232, 371, 375, 378
  of buffer, 302
electrokinetic separation, 302–303
  diffusion coefficient, 303
  on-chip, 303
electroluminescence, 505
electron, 538
  coherence length, 538
  density of states, 543
  gas, 542
  mean free path, 538
  mobility, 371, 378
  specific heat capacity, 542–544
  spin, 551
  thermal conductivity, 549
electron lithography, 559
electron–electron scattering, 137, 138
electron–phonon coupling, 140, 392, 507
electron–phonon scattering, 129, 137, 138
electrophoresis, 302–303, 334, 335
  analysis of PCR products, 305
  mobility, 302
  resolution, 303
  throughput, 303
ellipsometry, 134, 392
EM-LDOS, 459–463, 465
  resonance peak, 462, 463
embolus, 343, 348
emission line, 455, 507
  erbium, 509
  inhomogeneous broadening, 514
  intensity ratio, 328, 330, 332, 510, 511, 515, 516, 521–525
  of rare earth ion, 509
  temperature dependence, 511

thermal equilibrium, 509, 515
emissivity, 416, 419, 421, 428, 439
aluminium, 416
chromium, 434
lampblack, 416
measurement, 421
emittance, 415, 416
endothelial barrier, 346
enthalpy, 305
entropy, 305
entropy flow, 77, 81
maximal, 80–82
SWNT, 81
enzyme, 303
epitaxy, 373, 374, 383
equilibrium shift, 301, 305, 315, 320, 324, 327, 329, 334
erbium, 265, 509, 515, 527
Escherichia coli, 493
chemical mapping, 499, 501
electron microscope image, 494
PTIR study, 495
topographic image, 499, 501
ethylene glycol, 214
europium, 511
europium complexes, 312, 321
evanescent wave, 442–444
excimer formation, 322–323, 332
exciplex formation, 322–323
exciton, 514
exciton–plasmon coupling, 316
exposure, 415
extinction coefficient, 32, 128, 151, 198, 322, 470
by Mie theory, 199, 200
experimental, 199, 201
of nanoparticle ensemble, 162
spectrum, 198, 199, 201, 470
extinction cross-section, 163
by Mie theory, 165–167
of nanoparticle ensemble, 162
of oscillating dipole ensemble, 181
extinction efficiency, 160, 169
numerical calculation, 169
extinction index, 152
eye, 413

Fabry–Perot interferometry, 310
far field, 444
fast Fourier transform, 393
femtosecond laser, 333, 404, 507
Fermi energy, 392, 543, 550
Fermi liquid, 137, 138
Fermi temperature, 543
Fermi wavelength, 550
Fermi window, 232, 245
factor, 234
Fermi–Dirac distribution, 135, 137, 139, 542, 543
ferrimagnetic nanoparticle, 355, 358, 361
ferromagnetic nanoparticle, 355, 358, 361
Feynman diagram, 88
fibronectin, 345
finite-size effects, 139, 218, 538
flip chip, 390
fluctuation–dissipation theorem, 45
fluid–solid interface, 95
fluorescein, 326, 328, 331
fluorescence, 317, 506, 507
blue-edge, 330
contrast, 520
delayed, 323
intensity, 316, 317, 319, 321, 322, 513, 523, 528
lifetime, 317, 318, 321, 327, 333, 336, 509, 512, 513
modulated, 527, 528
pH dependence, 329
quantum yield, 316, 320, 327, 333, 506, 513
quenching, 326, 329, 333
spectrum, 319, 322, 329
temperature dependence, 319, 320, 508, 513
fluorescence correlation spectroscopy, 329
fluorescence imaging, 522, 523, 525
normalisation, 529–530
thermally modulated, 528–530, 532
fluorescence microscopy, 301
confocal, 336
intensity at given wavelength, 331
lifetime measurement, 333
procedures, 331–333
ratiometric intensity measurement, 332–333
two-photon, 336
fluorescence quantum yield, 315
fluorescent nanoprobe, 505, 515–517, 533
fluorescent probe, 311, 312, 336
acid–base reactions, 328–329
cyclodextrin complexation, 327–328
delayed fluorescence, 323
excimer formation, 322–323, 332
exciplex formation, 322–323
for pH, 332
isomerisation, 325
kinetic, 317–323
organic molecules, 319–320, 330
organometallic complexes, 320
population inversion temperature, 325
rare earth complexes, 321–322

ratiometric measurement, 332
response time, 325
sensitivity, 325
thermodynamic, 324–331
transition metal complexes, 320–321
fluorimetry, 309
fluorophore, 317
acid–base properties, 328–329
LMCT, 322
MLCT, 320
photodestruction, 331
TICT, 320, 328
focused ion beam, 3
folic acid, 347
force constant matrix, 65, 85, 88, 92
forensics, 304
four-bucket method, 394
Fourier law, 8, 17, 26, 52, 107, 114, 141, 142, 144, 146, 546
Fourier transform spectrometer, 471, 472
FRET, 326, 332, 336
FTIR spectroscopy, 471
by TRSTM, 463–465
of bacteria, 472, 473
of Candida albicans, 495
of Escherichia coli, 495
of PMMA film, 472, 473
of T5 virus, 496
fullerene, 312, 323
functionally graded material, 277
fungus, 492

gallium arsenide, 368, 373, 383, 384
carbon doped, 374
monocrystalline film, 377
on polycrystalline diamond, 375
thermal conductivity, 374
gallium nitride, 373, 374, 376, 384
monocrystalline film, 377
galvanometer mirror, 393
galvonometric scanning, 393
gas phase chromatography, 302
Gauss thermostat, 49
geometrical optics, 167
glass, 373, 508, 511
chalogenide, 446
doped, 515
fluoride, 515
substrate, 377
glioblastoma, 351
glycogen, 495, 498
GMS, 380
gold, 128–130, 134, 256, 371
band structure, 130
colloid, 163, 352
contact, 404
film, 454
nanoparticle, 142, 146, 351
nanosphere, 133, 448, 451
on silicon carbide, 458–463
plasmon resonance, 316
rod, 351
surface plasmon resonance, 131
thermo-optical response, 132
track, 402
grain boundary, 371, 375
graphene, 12, 55, 56, 77, 78
entropy flow, 81
thermal conductance, 78
graphite, 77, 542
dispersion relation, 79
entropy flow, 81
thermal conductance, 79
Green function, 63, 84
advanced, 89
equilibrium, 85–87
many-body, 84
non-equilibrium, 84, 87–89
resolvent, 65–67, 71, 72, 85, 86
retarded, 86, 89
single particle, 65
Green–Kubo relation, 45
GSM, 369
guarded hot plate, 46
gyrotron, 368

Hall effect, 266
Hankel functions, 164, 210
Harman method, 280
harmonic matrix, 65–67
Hashin–Shtrikman bounds, 217
heat current, 68–70, 83–84, 89–90
heat flux, 13, 108, 115, 142, 545, 546
volume average, 108
Heikes formula, 238
heparin, 346
heterodyne detection, 453, 454
heterodyning, 395
heterogeneous integration, 370
heterostructure integrated thermionic cooler, 246, 247, 261, 284
homodyne detection, 453, 454
human genome sequencing project, 303
hydrogen bond, 325, 334
hyperbolic diffusion equation, 144
hyperthermia, 350
magnetically induced, 350, 353–363
physiological constraints, 357–359

photoinduced, 351
hysteresis losses, 354, 357

illumination, 414
image contrast enhancement, 343, 348–350
image dipole method, 449–451
image processing, 412, 432–433
immune system, 344, 347, 493
immunoglobulin, 344
impedance matching, 379, 380
incandescence, 505
independent radiation–matter interaction regime, 162, 170–172
indium phosphide, 368, 373, 384
iron doped, 374
monocrystalline film, 377
thermal conductivity, 374
information transfer, 77, 81
infrared microscopy, 439, 457
numerical aperture, 441
resolution, 440
nanometric, 442
resolution limit, 442
infrared spectromicroscopy, 469–502
basics, 469–475
infrared spectroscopy, 439, 471–473
resolution, 439
integrated circuit, 302, 367, 373, 389, 390
active layer, 403
cooling, 261, 274, 390
encapsulation, 395, 399
flip chip, 390
imaging through substrate, 403, 405–407
monolithic UHF, 374
silicon, 403
temperature imaging, 395, 399, 402
integration, 302
density, 389
heterogeneous, 370
of calorimetry, 306
of light-emission functions, 370
of peristaltic pumps, 304
interatomic potential, 40, 92
interband transition, 128, 129, 133, 135, 136, 392
interconnects, 371, 390
interface liquid layer, 215
interface thermal resistance, 63, 109, 113, 114, 123, 140, 214, 216, 217, 547
intraband transition, 128, 129, 135
intratumoral injection, 348

Jablonski diagram, 506, 507
Joule effect, 227, 247, 256, 257, 261, 281, 303, 304, 307, 326, 332, 334, 335, 429, 434, 515, 521, 527, 551, 556, 559, 563

Kapitza resistance, 547
Keldysh method, 82
Kelvin relation, 227
kidney function, 349
kinetic theory of gases, 28, 56, 57
klystron, 368
Kramers–Kronig relations, 129, 465
Kubo formula, 63
Kupffer cell, 344

lab-on-a-chip, 301
for biological analysis, 301
heat transfer problems, 301–308
separation column, 303
separation time, 303
temperature servocontrol, 308
thermometric measurements, 331
Landau damping, 136
Landau theory, 137, 138
Landauer formula, 29, 72
lanthanides, 322
laser diode, 367, 371, 383, 506
cooling, 384
DFB, 383
lattice dynamics, 12, 63, 252–254
Hamiltonian, 82
harmonic, 65
laurdan, 316
fluorescence spectrum, 317
Legendre functions, 165, 206
light-emitting diode, 367, 371, 377, 505
blue, 373
energy efficiency, 383
green, 376
infrared, 373
near UV, 376
red, 373, 383
white, 505
Lindhard theory, 129
Liouville equation, 545
lipiodol, 360
liposome, 316, 350
Lippmann–Schwinger equation, 66, 71, 73
liquid crystal, 310–315
cholesteric, 309, 314–315
nematic, 309, 315
liver, 344, 345, 362
imaging, 349
LMCT state, 322
local energy conservation, 95, 99

local thermal equilibrium, 99–100, 102, 103
local thermal non-equilibrium, 105, 115
local thermodynamic equilibrium, 43
lock-in detection, 333, 389, 393, 399, 401, 403, 404, 452, 453, 458, 557, 562
Lorentz number, 550
Lorentz profile, 470
Lorenz number, 228, 233, 235
low temperature heat transfer, 22, 24, 42, 50, 537–565
  3D systems, 539
  interface thermal resistance, 547
  macroscopic, 539–544
  phonon speed, 20
  specific heat capacity, 546, 555–558
  steady state, 544–551
  surface effects, 546
  thermal conductance, 53, 548–550
  thermal conductivity, 59, 546
lowest critical solution temperature, 315, 352
luminance, 414
  temperature, 418
luminescence, 301, 324, 505–508
  effect of temperature, 507–508
  quantum dot, 513
  two-photon absorption, 507
luminescence spectroscopy, 318
luminous intensity, 415
lungs, 344
lymph system, 343, 349
lymphography, 349
lysosome, 345, 349

maghemite, 349, 359
magnetic resonance imaging, 348–350, 358
magnetite, 348, 349, 360
magnetron, 368
manganese perovskite, 358
matrix, 3, 107, 110, 140, 312, 330, 478, 480, 489, 509
  alumina, 141
  crystal field, 506
  crystalline, 512, 513
  dielectric, 140
  doped, 330
  Ge, 117, 119
  nanoporous, 197–204
  non-absorbing, 153, 156, 157, 164, 210
  optical index, 133
  oxide, 511
  oxysulfide, 511
  Si, 117
  silica, 141, 142, 197–204
  specific heat capacity, 141
  thermal conductivity, 108
  three-temperature model, 139
  with impurities, 114
  with LCST, 316
  Young's modulus, 490, 491
matrix–particle interface, 109, 123, 124, 143, 155, 210
Matthiessen's rule, 27, 129
Maxwell equations, 155, 206
Maxwell–Boltzmann velocity distribution, 43
Maxwell-Garnett theory, 132, 133
mediator, 353, 361
  capacitive, 354
  inductive, 354
  nanometric, 354
  surface functionality, 361
medical diagnosis, 304, 343, 344, 346, 351
membrane, 537, 541, 556, 560
  phospholipid, 316–317
  polyimide, 306
  self-supporting, 559
  silicon nitride, 306
  temperature profile, 557
  thermal conductance, 559
MESFET, 370
micelle, 350
microcapsule, 5
microchannel, 302, 334
microelectromechanical systems, 302, 370, 390
  temperature imaging, 434–435
microelectronics, 6, 17, 59, 301
  heat transfer problems, 367–385
  imaging, 505
  photothermal techniques, 389
  temperature imaging, 519
microfluidics, 301, 302, 306, 390, 533
  constriction, 332
  temperature measurement, 308–311, 313
    electrical methods, 308
    molecular probe, 310, 312
    optical methods, 309
microrefrigerator, 225, 256–263, 286
microwave oven, 368
Mie sequences, 164
Mie theory, 131, 163–175
  analytical solution, 205–211
  comparison with DDA, 191–197
  Du algorithm, 170
  extinction cross-section, 165–167
  limitations, 176
  MIEVO, 169
  radiative response of cloud, 173–175

radiative response of particle ensemble, 170–172
scattered field, 163
scattering cross-section, 165–167
scattering phase function, 165–167
military applications, 301, 369, 380
miniaturisation, 5, 302, 368
and innovation, 305
drawbacks, 307
in analytical sciences, 302
of calorimetry, 306
MLCT state, 320
molecular assembly, 316
molecular beacon, 313, 326, 332, 336
molecular beam epitaxy, 3, 255
molecular bonding, 374, 375
molecular dynamics, 12, 21, 40–49, 63, 64, 92, 118
equilibrium, 43, 45, 56
for nanofluid, 215
homogeneous non-equilibrium, 43, 49, 56
limitations, 41
non-equilibrium, 56, 57
non-homogeneous non-equilibrium, 43, 46–48
molecular probe, 301, 308–310, 312, 313
sensitivity, 313
molecular recognition, 346, 347
molecular thermometer, 308, 315, 320, 323, 326, 332
from biological entity, 327
luminescent, 311
rare earth complexes, 321
molecular vibration, 439, 470, 471
mononuclear phagocyte system, 344, 345
Monte Carlo simulation, 10, 27, 33–40, 50, 235–236
of Peltier coefficient, 235, 286
of Seebeck coefficient, 235
Moore's law, 381
Mott–Anderson transition, 553
mouse, 351
multiplexing, 391, 395
multispectral measurements, 427–431
non-ideal, 427–428
three wavelength, 428–429
mycosis, 492

NAIL, 405
nanoaperture, 440, 445–447
controlling position, 446, 447
in water, 447
optical fibre, 446
nanocalorimeter, 306, 309, 390, 557
heat capacity, 306
response time, 306
thermal conductance, 306
nanocomposite, 58
athermal regime, 137, 139
heat conduction in, 95–105, 107–125
heat exchange dynamics, 136–147
magnetic microparticle, 359
metal–dielectric, 127–147
optical response, 127–147
infrared, 440
particle shell model, 110–113
particulate, 151–211
phonon scattering in, 114–121
pulsed laser excitation, 134, 135
radiative properties, 151–211
thermal regime, 139–147
thermo-optical response, 132–136
two-temperature model, 139
ultrafast excitation, 142
nanoconvection, 214, 215
nanoelectromechanical system, 370
nanofilm, 3, 17, 29, 31, 33, 50–54
argon, 41
germanium, 50
low temperature, 50
metal, 59
Pt, 59
silica, 59
silicon, 41, 50–53, 58
temperature profile, 50
thermal conductivity, 51–54, 58, 59
nanofluid, 17, 58
thermal conductivity, 213–219
nanomaterial, 3
mid-infrared s-SNOM, 454
nanomedicine, 343–363
definition, 343
nanoparticle, 6, 17, 343
absorbed power, 156, 157, 160–161
absorption by, 153
absorption cross-section, 160–161
absorption efficiency, 160
aggregation, 152, 213, 217, 218, 348
as mediator, 354
cluster, 200, 215, 217, 218
colloidal suspension, 213, 356, 359
copper, 213
effect on thermal conductivity, 121
encapsulated, 360
ensemble, 162
radiative response, 170–172
ErAs, 121, 265–269

extinction cross-section, 131, 159–161, 165–167
extinction efficiency, 160
extinction power, 157–161
ferrimagnetic, 355, 358, 361
ferromagnetic, 355, 358, 361
gold, 142, 146, 351, 448, 451
hydrodynamic volume, 349, 356
hydrophilic, 346
hydrophobic, 345
in blood, 344, 345
in vivo applications, 344–352
KBr, 120
KCl, 120
lead fluoride, 511, 515
maghemite, 349, 359
magnetic, 349, 351, 352, 354
magnetite, 349, 360
metal, 127–147, 351
noble metal, 128–132
optical response, 131
PbSe, 4
photoluminescence spectrum, 511
plasma half-life, 345, 346
powder, 330
representative cluster, 201, 202
scattered power, 156–161
scattering by, 153, 154
scattering cross-section, 159–161, 165–167
scattering efficiency, 160
scattering phase function, 160, 165–167
semiconductor, 508
Si, 119
silica, 5, 197, 199
silver, 120
superparamagnetic, 349, 355, 357, 359, 360
surface plasmon resonance, 131
thermal conductivity, 41
vascular lifetime, 345
nanostructure, 3, 17
heat conduction in, 8
thermal conductivity, 58
nanotube, 4, 14, 17, 29
armchair, 78
carbon, 41, 44, 54–57, 77, 80, 213, 216, 217, 542, 561, 562
entropy flow, 80, 81
for cooling, 6
heat flux in, 7, 9
MWNT, 59
specfic heat capacity, 542
suspended, 8
SWNT, 59
thermal conductance, 77, 561
thermal conductivity, 3, 59
transmission function, 77
zigzag, 78
nanowire, 6, 8, 12, 17, 22, 28, 29, 31, 38, 225, 281, 537, 548
analytic model, 30–31
array, 243
as acoustic waveguide, 29
composite, 242
metal, 527
PbTe, 230
rough, 230, 255
Seebeck coefficient, 255, 282, 561
silicon, 4, 12, 24, 34, 54, 58, 91, 230, 561, 562, 564, 565
surface effects, 546
suspended, 562
thermal conductance, 560, 561, 565
thermal conductivity, 3, 8, 15, 31, 58, 255, 547, 562–564
NBD, 316
near field, 442–444, 452
definition, 444
detection, 444, 448, 452–454
heat transfer, 140, 216, 463
thermal radiation spectrum, 465
near-field microscopy, 440, 444–456, 475, 529
heat sink effect, 411
mid-infrared, 446
physical principle, 444–445
resolution, 440
near-field probe, 444, 505
Néel relaxation, 355–357
nematic liquid crystal, 309
neutron scattering, 20
Newton's second law, 40
nickel strip, 525–527, 532
topographic image, 528, 529
noble metal
nanoparticle, 128–132
specific heat capacity, 542–543
thermo-optical response, 132
noise equivalent power, 422
non-Fourier effect, 13
confinement, 10, 11
dimensionality, 12, 13
rarefaction, 9
non-radiative recombination, 506, 507, 512
non-steady state conditions, 34, 50, 101, 380
nonlinear optical susceptibility, 136
nonlinear optics, 131, 135
normal distribution, 419, 420
normal process, *see* phonon scattering
nuclear fusion, 368

nuclear magnetic resonance, 308, 309, 313, 334–335
nucleic acid, 326–327
numerical aperture, 396, 399, 402, 405–407, 425, 441
  mirror objective, 441
  silica objective, 441

oligonucleotide, 327
Onsager reciprocity relation, 227
opsonin, 345
opsonisation, 344, 345, 349
optical fibre, 330, 440
  AgClBr, 446, 447
  chalcogenide glass, 446
  silica, 446, 447
optical index, 128, 151, 152, 469–471, 475
  contrast, 164, 169, 211
  position dependent, 178
optical theorem, 157, 158
optoelectronic components, 377, 440
  mid-infrared s-SNOM, 454
optronics, 384
organic dyes, 320, 330
  band spectrum, 331, 332
organometallic complex, 320
oscillating dipole, 176

P1 method, 144
parallelisation, 302
particulate medium, 152
passivation, 371, 400, 434, 521
  layer, 390, 395, 401, 402
passive targeting, 345
pathogen detection, 304
Pauli exclusion principle, 129, 138, 542
PEG, *see* poly(ethylene glycol)
pegylation, 346
Peltier coefficient, 227, 232, 242, 247
  bias-dependent, 247, 285
  Monte Carlo simulation, 235, 286
  of InGaAs barrier, 247
Peltier effect, 226, 426
  for cooling, 227, 274, 285
  in bulk silicon, 256
  nonlinear, 247, 286
Peltier element, 305, 399, 403
peptide, 327, 347
peristaltic pump, 304
pH probe, 332, 336
phase speed, 20
phase transition, 301, 311
  at small length scales, 538
  gel to liquid, 316
phase-change material, 5
phenol red, 328
phonon, 7, 18–27, 507, 538
  absorption, 138
  angular frequency, 143
  anharmonic transport, 82–91
  black body, 32, 38, 541, 547
  collisions with boundaries, 24, 25, 116
  collisions with defects, 24, 25, 39, 115
  collisions with electrons, 129, 137, 138
  collisions with impurities, 116
  collisions with particles, 116
  collisions with phonons, 25, 26, 39, 115
  confinement, 24, 28, 50–56, 237, 253, 538
  coupling with electrons, 140
  coupling with excitons, 514
  creation operator, 83
  current, 90–91
  density of states, 12, 22, 143, 546
  destruction operator, 83
  diffuse reflection, 9, 28, 32, 33, 38, 54
  dispersion, 12, 28, 77, 237, 540, 542
  equilibrium intensity, 143
  frequency spectrum, 40, 90
  gas, 542
  group velocity, 20, 28, 37, 143, 237
  harmonic transport, 92
  intensity, 32, 33, 143
  lifetime, 141
  longitudinal, 22, 37, 117
  mean free path, 10, 17, 24, 25, 63, 107, 114, 137, 141, 142, 144, 538, 546
  nanoscale transfer, 548
  number, 23, 35–37, 41–43, 70, 86
  phase speed, 20
  polarisation, 22, 26, 35, 36
  polariton, 459
  quantisation, 21, 57, 83
  sampling, 35–39
  scattering, 24–26, 28, 39, 89, 90, 547, 561
    by impurities, 115
    by nanoparticles, 118
    by particles, 117–118, 125
    in nanocomposite, 114–121
    normal process, 25, 26, 39
    umklapp process, 25, 26, 39, 116, 119
  specific heat capacity, 539–542
  spectral intensity, 23
  spectrum, 116, 118, 139, 538, 541, 542, 548
  specular reflection, 9, 10, 28, 32, 33, 38, 547, 561
  spontaneous emission, 138
  stimulated emission, 138
  thermal conductance, 546–549

thermal conductivity, 546–549
transport, 63, 64, 66, 237
transverse, 22, 37
tunneling, 215
phonon glass/electron crystal material, 229
phospholipid bilayer, 350
phospholipid membrane, 316–317
phosphorescence, 317, 323, 333, 506, 507
lifetime, 327, 509
phosphorescent probe, 327
photobleaching, 508
photoluminescence spectrum, 316, 330, 510
temperature dependence, 511, 513, 514
photoluminescence techniques, 391
photomultiplier, 418, 423–424, 426
biasing, 423
gain, 423
pass band, 424
quantum efficiency, 420, 426
spectral sensitivity, 426
photon counting, 419, 425, 429
photon flux, 418
randomness, 419–420
ratio, 428
temperature dependence, 419, 429
photophysical phenomena, 311, 317–319, 324
fluorescence, 317
internal conversion, 317, 318
intersystem crossing, 317, 318, 321, 323
phosphorescence, 317
radiative relaxation, 318
photopic vision, 413
photothermal effect, 469, 476, 477, 487, 489
photothermal induced resonance technique, *see* PTIR
photothermal techniques, 389–407
PTIR, *see* PTIR
photothermal therapy, 351
pig, 359
pixel, 393
pixel addressing, 372
Planck distribution, 21, 39, 41, 42, 415–417, 429, 439, 457, 541
UV–visible emission, 418
plasma deposition, 3
plasma frequency, 392
plasma half-life, 345, 346
of paramagnetic ions, 349
SPIO, 349
stealth particle, 346
USPIO, 349
plasmon, 128, 136, 316, 459
plasmon resonance, 316
plasmonics, 127
mid-infrared s-SNOM, 454
platinum
complexes, 321
thermometer, 560
PMMA, 473, 478
absorption by, 488
thermal diffusivity, 479
Poisson distribution, 419, 420
polarisability, 187–191
Clausius–Mossotti model, 187–188, 192
Doyle model, 190
Draine model, 189–190, 192, 195
polariton resonance, 455
poly(ethylene glycol), 316, 346
polyamide, 396, 400
polyimide membrane, 306
polymer capsule, 350
polymerase chain reaction, 303–305
polysaccharide, 346
polysialic acid, 346
polysilicon circuit, 521–522
polysilicon resistance, 396, 400
temperature image, 401
polysilicon strip, 522
porosity, 97, 197
porous medium, 95, 104
power generation, 225–229, 245, 246, 251, 272–274, 276, 277, 282
Poynting vector, 154–156
predistortion, 379
prostate cancer, 351, 361
protein, 329
protonation, 325, 328, 329, 333
PTIR, 475–477
basic idea, 476
detection sensitivity, 488
experimental results, 492–501
lateral resolution, 488–491
of Candida albicans, 495
of Escherichia coli, 495
of T5 virus, 496
setup, 475
pump–probe experiment, 50, 131, 135, 136
pyrene, 316

quantum cascade laser, 455, 456
quantum dot, 225, 243, 248, 251, 255, 316, 508, 513–514
PbSeTe, 7
quantum efficiency, 422, 506
photomultiplier, 420, 426
quantum photodetector, 422
dynamic range, 422
NEP, 422

sensitivity, 422
specific detectivity, 422
spectral band, 422
threshold, 422
quantum well, 228, 234, 241–243
Bi, 242
PbTe/PbEuTe, 241
Si/SiGe, 241
quantum wire, 228, 241–243, 285
Bi, 242
thermal conductivity, 255
quantum yield, 316, 318, 320, 327, 333, 506
quasi-steady state, 101
quenching, 326, 327, 329, 333

rabbit, 359
radiative flux, 412, 419
radiative lifetime, 317, 318
radiative recombination, 509
radiative transfer, 160
in nanoporous silica, 197–204
radiative transfer equation, 23, 31, 151, 198
steady state, 151
radiometry, 411
beam throughput, 413–414
CCD camera, 430–435
calibration, 433
dynamic detection, 431
image processing, 432–433
optimisation, 433
static detection, 431
dynamic detection, 431
emissivity, 419, 421
emittance, 415, 416
error in temperature, 421
exposure, 415
illumination, 414
imaging MEMS, 434–435
luminance, 414
luminance temperature, 418
luminous intensity, 415
microscope transfer function, 420, 421, 426–429
multispectral measurements, 420–421, 426–431
non-ideal, 427–428
three wavelength, 428–429
photomultiplier, 423–424
biasing, 423
gain, 423
pass band, 424
photon flux, 418
randomness, 419–420
temperature dependence, 419
physical principles, 412–415
quantum photodetector, 422
dynamic range, 422
NEP, 422
sensitivity, 422
specific detectivity, 422
spectral band, 422
threshold, 422
radiative flux, 412, 419
relevant wavelengths, 412
sensitivity, 413, 422, 431
short wavelength, 411, 417–419, 421–435
spatial resolution, 422, 430
static detection, 431
temperature resolution, 422
UV thermal microscope, 425–430, 436
calibration, 429
experimental results, 429
radiotherapy, 350, 353
Raman scattering, 334
Raman spectroscopy, 308, 309, 313, 331, 334, 391
rare earth complexes, 321–322
rare earth ion, 330, 332, 507–511
optical transition, 509
trivalent, 509
ratiometric measurement, 317, 326, 328, 330, 334, 336, 510, 511, 515, 516, 521–525
Rayleigh criterion, 411, 417, 439, 441, 442, 457, 490
Rayleigh scattering, 167–169, 334
Rayleigh–Gans scattering, 167
reaction rate, 301
temperature dependence, 317
ultrafast, 307
recursion method, 67
red blood cell, 346
reflection coefficient, 392, 395
temperature dependence, 395, 401
refractive index, 128, 151, 152, 310, 470
silicon, 406
relaxation time, 25, 28, 115, 116, 119, 121, 144, 324, 325, 479, 480
approximation, 23, 30, 31, 39, 137, 138, 142, 545
Brown, 356
collisional, 22, 23, 27
electron, 392
Néel, 356
renal clearance, 350
resistive thermal detector, 306, 309
resolution limit, 441–444
infrared microscopy, 442
rhodamine 3B, 312

rhodamine 3G, 319
rhodamine B, 312, 331, 508
  lactone form, 319, 320
  zwitterion form, 319, 320
Ricatti–Bessel functions, 164, 170, 211
Richardson equation, 247
rock wool, 5
Rosei model, 129, 130, 136
ruthenium complexes, 312, 320, 321, 331

sapphire, 373
scanning electron microscopy, 217, 218
scanning thermal microscopy, 411, 505
  a.c. measurements, 527–531
  applications, 519–533
  d.c. measurements, 520–526
  experimental setup, 518–519
  lateral resolution, 515, 517, 519, 524, 525
  tip–sample heat transfer, 531–533
  with thermocouple, 533
  with thermoresistive wire, 533
scanning tunneling microscope, 440, 445, 464
  TRSTM, *see* TRSTM
scattering albedo, 151, 162, 198
  by Mie theory, 199, 200
  experimental, 200, 202
  spectrum, 199, 200, 202
scattering coefficient, 151, 211
  numerical calculation, 169
  of nanoparticle ensemble, 162
scattering cross-section, 163, 195
  by Mie theory, 165–167
  of gold nanosphere, 451
  of oscillating dipole ensemble, 181
scattering efficiency, 160, 169
  numerical calculation, 169
scattering phase function, 151, 160, 163, 194, 197
  by Mie theory, 165–167
  of nanoparticle ensemble, 162
  of oscillating dipole ensemble, 183
Schottky anomaly, 544
Schottky junction, 306
Schottky rectifier, 370, 376
scotopic vision, 413
screening, 305
Seebeck coefficient, 225, 227, 246, 257, 267–269, 275, 277, 278, 280
  at low temperature, 243
  bias-dependent, 285
  calculation, 232
  cross-plane, 234, 249, 266, 271, 279, 280
  enhancement, 229, 239, 264
  measurement, 279, 281
  Monte Carlo simulation, 235
  of GaAs, 233
  of multibarrier structure, 247
  of nanowire, 255, 282
  of quantum dot array, 243
  of strongly correlated oxide system, 238
  of superlattice, 234, 279, 280
  spin, 238
  trade-off with electrical conductivity, 239–240, 248, 249
Seebeck effect, 225, 227, 552, 561
  at low temperature, 552
  power generation, 227
self-energy matrix, 87, 88
semi-transparent medium, 151
  heterogeneous, 152
semiconductor, 378, 380
  active components, 369
  amorphous, 371
  band gap, 371, 507, 514
    temperature dependence, 392
  components, 370
  direct band gap, 383
  doping, 372
  grain boundary, 371, 375
  heat conduction in, 370
  III–N, 373, 374, 376, 379, 380
  III–V, 367, 377
  microscopic order, 371–372
  mid-infrared s-SNOM, 454
  monocrystalline, 371
  nanoparticle, 508
  photoluminescence spectrum, 514
  polycrystalline, 371
  thermal conductivity, 371, 372
  wide band gap, 368, 378
Si–Ge interface, 91
signal processing, 371, 412
  SNOM, 452–454
  TRSTM, 460, 461, 464
silica, 134, 330, 371, 384, 397
  fumed, 197, 198
  in microelectronics, 398
  insulating layer, 377
  matrix, 141, 142, 197–204
  mesoporous particles, 352
  nanoparticle, 199
  optical fibre, 446, 447
  optical transmission, 446
silicon, 305, 367, 370, 375, 384, 403, 405, 413
  enriched, 372
  for CCD, 425
  imaging through, 403, 405–407
  membrane, 557, 558

nanowire, 58, 230, 561, 562, 564, 565
on diamond, 374
Peltier effect in, 256
photosensitivity, 425
polycrystalline, 372
reflectivity, 403
refractive index, 406
substrate, 391, 404, 434
wafer, 373
silicon carbide, 375–376, 379, 384
doped, 376
growth, 376
polariton resonance, 455
polycrystalline, 372, 376
substrate, 373, 455, 458–463
thermal conductivity, 376
titanium doped, 374
vanadium doped, 374
wafer, 376
silicon nitride, 371, 396
absorption by, 401
membrane, 306
temperature sensor, 557, 559
thermal conductivity, 377
silicon on insulator, 279, 381, 390
silver, 128, 371
band structure, 130
surface plasmon resonance, 131
smoke, 152
SNARF, 328
SNOM, 440
active fluorescent tip, 445, 456
aperture, 445–448
mid-infrared, 446, 447
apertureless, 440, 448–456
heterodyne detection, 453, 454
homodyne detection, 453, 454
infrared, 454–456
physical principle, 448–449
resolution, 440, 448
signal processing, 452–454
biological samples, 447, 454
dielectric contrast, 449–451, 454
essential elements, 457
image, 444
image dipole method, 449–451
nanomaterials, 454
resolution, 440
scattering, 440, 448–456
sub-wavelength resolution, 444, 448
solid angle, 413
solid immersion lens, 405–407
solid–liquid interface, 96
soliton, 7
sound speed, 20, 539
specific absorption rate, 354
specific detectivity, 422, 423
of photomultiplier, 424
specific heat capacity, 22, 116, 144, 477, 550
effect of phonon confinement, 538
electrical measurement, 551–564
in Debye model, 541
in low-dimensional systems, 541–542
low temperature, 555–558
of electrons, 542–544
of matrix, 141
of metal, 141
of MWNT, 542
of nanocalorimeter, 306
of noble metal, 542–543
of phonons, 539–542
of two-level system, 543–544
relation with thermal conductivity, 546
temperature dependence, 537, 540–543, 546
specific intensity, 151
specific loss power, 354, 355, 357, 358, 361
spectrophotometry, 309
specularity, 10
spherical harmonics
scalar, 205
vector, 164, 206
SPIO, 349, 357, 360
plasma half-life, 349
spleen, 344, 345, 362
stealth particle, 344, 345
plasma half-life, 346
Stefan distribution, 417
steric barrier, 346
Stokes excitation, 331, 334, 509
sub-wavelength resolution, 475
submicron strip, 525
sugar, 347
superinsulating materials, 197
radiative transfer properties, 198
superlattice, 6, 58, 225, 236, 239, 242, 243, 246, 247, 256, 270, 278, 285
as momentum filter, 248
ballistic transport in, 236
BiTe, 281
BiTe/PbTe, 261
BiTe/SeTe, 280
design, 236
ErAs/InGaAs, 265–267
ErAs/InGaAs/InGaAlAs, 281
figure of merit, 280
for cooling, 256, 258, 260–263
GaAs/AlAs, 253
III–V, 234

InGaAs/InAlAs, 234
InGaAs/InGaAlAs, 234, 249
metal-based, 247, 271
metal/semiconductor, 270
metallic, 264
miniband, 234
narrow-band, 234
of ErAs islands, 265
on SOI, 279
PbTe, 241
period, 233, 254
phonon modes, 253
ScN/(Zr,W)N, 270
Seebeck coefficient, 234, 279, 280
short-period, 266
Si/Ge, 253
Si/SiGe, 4
SiGe, 260, 262, 263
SiGe/Si, 261–263
thermal conductivity, 252–254, 278
thermal diffusivity, 279
thick-period, 253
superparamagnetism, 349, 361
surface plasmon resonance, 131
for hyperthermia, 351
in copper, 131
in gold, 131
in silver, 131
surgical ablation, 350
suspended membrane nanocalorimeter, 557
switching phase, 378, 379
synchrotron, 368
syphilitic paralysis, 353

T5 virus, 493, 501, 502
chemical mapping, 500
PTIR study, 496
topographic image, 500
TAMRA, 328
Taylor–Aris dispersion, 303
technological bottleneck, 302
telecommunications
amplifier, 379, 380
base station, 368, 369, 379, 380
rising frequencies, 369
satellite, 368, 369
temperature deviation, 98, 100–102
temperature field, 96, 112, 113, 135, 151
measurement, 308
temperature gradient, 17, 24, 46, 107, 108, 112, 115, 544–546, 559
continuous, 557, 559
time-varying, 559, 561
temperature imaging, 439
of aluminium oxide strip, 522–524
of lab-on-a-chip, 301
of microelectromechanical systems, 434–435
of microfluidics, 311
of polysilicon circuit, 521
of resistor, 326
of submicron strip, 525
remote, 308
rhodamine B, 319
with alternating current, 527–529
with direct current, 520–526
temperature modulation, 389, 390, 399, 400, 555–558, 561–564
temperature probe, 511
terbium complexes, 321
Texas Red, 326
therapy, 343, 344, 346, 350, 351
thermal bystander effect, 360
thermal conductance, 33, 52, 70, 77, 544, 547
3ω method, 561–564
at interface, 110, 114
ballistic, 80
cross-plane, 52, 53
dynamical measurement, 561–564
effect of phonon confinement, 538
electrical measurement, 551–564
low temperature, 75–76, 80, 548–550
maximal, 77–80
nanoscale, 559–564
of carbon nanotube, 561
of disordered solid–solid interface, 91
of graphene, 78, 80
of graphite, 79, 80
of membrane, 559
of membrane support, 560
of nanocalorimeter, 306
of nanofilm, 51
of nanowire, 560, 561, 565
of SWNT, 78, 80
phonon, 546–549
steady-state measurement, 559–561
surface effects, 546
temperature dependence, 549
thermal conductivity, 14, 15, 17, 24, 38, 116, 144, 227, 232, 480, 546
bulk, 63
by molecular dynamics, 41, 43–49
compared with electrical conductivity, 550
cross-plane, 33, 51, 563
effect of doping, 120, 121, 372
effect of nanoparticles, 121
effective, 104, 107, 110
electron, 549

in composite, 107, 113, 114, 118, 119
in Debye approximation, 22
in disordered alloy, 120
in two-phase system, 104
in-plane, 33, 51, 53, 59
kinetic method, 544–545
low temperature, 59, 119, 546
low-dimensional limit, 550
non-Fourier effect, 13
of amorphous solid, 63
of buffer, 302
of diamond, 374
of dielectric crystal, 27, 28
of electrical insulator, 371
of GaAs, 374
of InP, 374
of matrix, 141
of metal, 228
of MWNT, 59
of nanofilm, 51–54, 58, 59
of nanofluid, 213–219
of nanoparticles, 41
of nanostructure, 28–29, 58
of nanotube, 3, 54–57, 59
of nanowire, 3, 8, 15, 31, 54, 55, 58, 255, 547, 562–564
of quantum wire, 255
of semiconductor, 27, 228, 371, 372
of SiC, 376
of SiN, 377
of superlattice, 252–254, 278
of SWNT, 59, 77
phonon, 237, 546–549
reduction, 225, 228–230, 253, 254
temperature dependence, 537, 546
thermal diffusivity, 215, 397, 479, 480
of superlattice, 279
thermal diode, 7
thermal expansion, 375, 469, 478, 481
AFM detection, 482–485
image, 488, 489
thermal insulation, 5, 197
thermal lensing, 135
thermal nanosensor, 505
thermal noise, 488, 562
thermal runaway, 378
thermo-optical response, 132–136, 395, 398, 404
nanocomposite, 133–134
noble metals, 132
steady state, 135
transient, 135
thermoablation
clinical trials, 359
selective, 353
total, 353, 358
thermochromic probe, 310, 314, 329
thermocouple, 227, 306, 309, 310, 429, 533, 552
Thermocule, 6
thermoelastic coefficient, 477, 481
of organic material, 489
thermoelastic deformation, 481, 486
thermoelectric conversion, 6
thermoelectric effect, 225–287, 552
coefficient of performance, 228
figure of merit, 225, 227, 238, 242, 262, 267, 269
functionally graded materials, 277
nonlinear, 285
on nanoscale, 561
thermographic phosphors, 330
thermography
by CCD camera, 431
calibration, 433
camera, 436
infrared, 308, 439
of microsystems, 322
short wavelength, 421–435
thermometry, 331, 508, 537
concentration-independent, 317
electrical, 551–554
for microfluidics, 308–311, 313
electrical methods, 308
molecular probe, 310, 312
optical methods, 309
four-wire setup, 309, 551, 552, 560
invasive, 308
low temperature, 551–554
luminescent materials used for, 508–514
molecular tagging, 336
non-invasive, 308
on-chip, 332
optical, 389, 391, 411, 510
organic molecules used for, 319–320, 508–509
Raman techniques, 334
resistive, 551, 553
semi-invasive, 308
sensitivity, 309
spatial resolution, 309
temporal resolution, 309
using fluorescence, 411
using radiometry, 411
thermomodulation experiments, 132, 135
thermophoretic velocity, 215
thermopile, 306, 309
thermoradiotherapy, 353

thermoreflectance imaging, 58, 391–407, 411
  CCD setup, 390
  four-bucket method, 394
  near-IR, 391, 403–407
  physical phenomenon, 392
  resolution, 391, 402, 405–407
  UV, 391, 401–402
  visible, 391, 395–400
thermoreflectance spectroscopy, 395–398
  calibration, 398–400
  modelling, 396–398
  time-domain, 279
thermoresistive wire, 533
thermosensitive materials, 311–317
  hydrogel polymer, 352
  liquid crystal, 311–315
  phospholipid membrane, 316–317
  polymer, 315–316
thermotherapy, 361, 362
  by implanted heat source, 353
  contact, 353
  non-contact, 353
  SLP, 358
  ultrasound, 358
thermotunneling, 244
3ω method, 561–564
three-temperature model, 139–147
throughput, 413–414
TICT state, 319, 320, 328
time gate, 404, 405
titanium, 404
TMapper, 400
tobacco virus, 475
tomography, 361
topographic image, 489
  of aluminium strip, 523
  of E. coli, 499, 501
  of nickel strip, 528, 529
  of polysilicon strip, 522
  of T5 virus, 500
topography, 475
toxicity, 331, 350, 352, 360
transfer function, 420, 421, 426–429
transformer, 369
transistor, 367, 368, 370, 376, 379, 389
  capacitance reduction, 381
  cooling, 377
  elementary operation, 381
  energy recovery, 382
  in mobile phone, 379
  UHF, 376
  voltage reduction, 382
transition ion, 512–513
  fluorescence, 513
transition metal complexes, 311, 320–321
  absorption, 329
transmission function, 70–75
  Landauer formula, 72
  of nanotube, 77
  of SWNT, 79
  three-region formula, 71
  two-region formula, 72
  weak coupling, 76
transmittance, 470
transmittivity, 420, 426–428
travelling wave tube, 368
TRSTM, 440, 456–465
  applications, 458–463
  approach–retract curve, 461
  as temperature probe, 463
  contrast inversion, 462
  FTIR spectroscopy, 463–465
  resolution, 440, 463
  sensitivity, 463
  setup, 457–458
  signal processing, 460, 461, 464
  temperature dependence of signal, 464
TTA, 511
tumour, 343, 346, 348, 353, 359, 362
  active targeting, 346
  deep, 354
  treatment by superparamagnetic nanoparticles, 360
  well-perfused, 358
tungsten, 517
two-phase system, 95, 96
  thermal conductivity, 104
two-temperature model, 139

UHF components, 376, 377, 379
  in mobile phones, 383
ultralocalised IR spectroscopy, 494–496
umklapp process, *see* phonon scattering
UMTS network, 369, 380
up-conversion, 507
USPIO, 349, 357, 360
  plasma half-life, 349
UV thermal microscope, 425–430, 436
  calibration, 429
  experimental results, 429
  multispectral, 426

Van der Pauw configuration, 279, 399
Van't Hoff law, 307, 325, 334
Varshni's empirical law, 514
vascular lifetime, 345
vertical cavity surface emitting laser, 273
vesicle, 317

virus, 493
  tobacco, 475
VLSI, 367, 373
Völklein method, 559, 560
volume average
  intrinsic, 97
  of heat flux, 108
  superficial, 97
volume averaging, 96–105
  closure problem, 103–104
  periodic volume, 102

wafer, 372
  diamond, 374, 375
  SiC, 376
WDM optical communication network, 273
Weiss domain, 354
Wiedemann–Franz law, 228, 251, 264, 550
Wien's law, 416, 417
WiFi, 380
Wimax, 369, 380
work function, 422
wurtzite, 376

yeast, 492
Young's modulus, 482, 490, 491
ytterbium, 509, 515

Zener diode, 423
zero point energy, 539